Capítulo 4

Análise de Dados Privada em Redes Sociais

André L. C. Mendonça, Felipe T. Brito, Javam C. Machado

Abstract

With the increasing use of social networks, analyzing user interactions has become crucial for understanding social dynamics and discovering valuable insights under a variety of domains. However, in the context of social network analytics, concerns regarding individuals' privacy persist, requiring measures to protect sensitive information. In recent years, differential privacy has become the de facto standard for privacy-preserving data analysis under strong mathematical guarantees based on the concept of indistinguishability. This chapter provides a comprehensive overview of existing differentially private methods and techniques to protect sensitive information while enabling meaningful social network analysis. We explore the principles of differential privacy, highlighting its mechanisms for adding noise to data to prevent individual re-identification. Additionally, we investigate the strategies for applying differential privacy in social network analytics, encompassing data publishing, graph analysis, and machine learning tasks privately.

Resumo

Com o uso crescente das redes sociais, a análise das interações dos usuários tornou-se crucial para compreender as dinâmicas sociais e descobrir informações valiosas em diversos domínios. No entanto, no contexto de análise de dados em redes sociais, existem preocupações inerentes à privacidade dos indivíduos, as quais exigem medidas para proteger informações sensíveis. Nos últimos anos, a privacidade diferencial tornou-se o padrão para prover privacidade na análise de dados sob fortes garantias matemáticas baseada no princípio da indistinguibilidade. Este capítulo fornece uma visão geral dos métodos e técnicas diferencialmente privadas para proteger informações sensíveis e, simultaneamente, permitir análises relevantes de redes sociais. Exploramos os princípios da privacidade diferencial, destacando seus mecanismos para adicionar ruído aos dados para evitar a reidentificação dos indivíduos. Além disso, investigamos as estratégias para aplicar privacidade diferencial na análise de dados em redes sociais, abrangendo a publicação de dados, a análise de grafos e tarefas de aprendizado de máquina de maneira privada.

4.1. Introdução

Redes sociais são plataformas *online* que possibilitam a interação entre pessoas por meio do compartilhamento de informações, ideias, interesses e outros tipos de conteúdo. Essas plataformas, que incluem exemplos amplamente conhecidos como Facebook, Twitter, Instagram e LinkedIn, funcionam como sistemas que conectam usuários com base em suas relações pessoais, profissionais ou interesses comuns. A evolução das redes sociais começou no final dos anos 90 e início dos anos 2000 com sites como Six Degrees e Friendster [Boyd and Ellison 2007], que permitiam aos usuários criar perfis e conectar-se com amigos. Com o lançamento do MySpace e, posteriormente, do Facebook, as redes sociais começaram a se expandir rapidamente, oferecendo recursos mais avançados como compartilhamento de fotos, vídeos e eventos, além de mecanismos de comunicação direta, como mensagens instantâneas e comentários.

Hoje, as redes sociais desempenham um papel central na vida moderna, influenciando a maneira como as pessoas se comunicam, consomem informações e realizam negócios. Elas oferecem um espaço virtual onde indivíduos podem expressar suas opiniões, manter contato com amigos e familiares, seguir notícias e tendências, e até mesmo buscar oportunidades de emprego. Empresas e organizações também utilizam essas plataformas para marketing, atendimento ao cliente e construção de marca. Além disso, as redes sociais têm um impacto significativo na formação de comunidades e movimentos sociais. Elas facilitam a organização e mobilização de grupos em torno de causas comuns, permitindo que indivíduos compartilhem suas experiências e apoiem uns aos outros de maneira rápida e eficaz [Prell 2011].

Como reflexo da vida social real, as redes sociais fornecem um meio para compartilhar muitas informações privadas e sensíveis. Por exemplo, em muitas redes sociais online, é comum que os usuários sejam solicitados a fornecer informações pessoais, como nome, gênero, data de nascimento, nível de educação, estado civil, foto pessoal e até mesmo número de telefone celular. Além disso, conteúdos gerados pelos usuários, como textos, fotos, vídeos e localizações geográficas, também são armazenados em bancos de dados [Baden et al. 2009]. Esses dados podem ser compartilhados com terceiros para serviços comerciais adicionais, como análise de dados, publicidade direcionada, recomendações e avaliações de aplicativos. Contudo, uma vez que dados de redes sociais contêm informações sensíveis, compartilhar esse tipo de dado sem garantias suficientes de privacidade pode comprometer seriamente a privacidade dos indivíduos [Brito et al. 2024]. As leis e regulamentações atuais sobre privacidade de dados, como o *General Data Protection Regulation (GDPR)* [European Commission 2018] e a *Lei Geral de Proteção de Dados Pessoais (LGPD)* [Brazil 2018], exigem que os indivíduos não possam ser reidentificados a partir das informações divulgadas.

Se informações pessoais privadas forem acessadas por terceiros que, teoricamente, não deveriam ter acesso a elas, os indivíduos afetados podem enfrentar várias consequências negativas. Isso inclui a exposição a ataques de intrusão, como *spam*, mensagens indesejadas e, em casos mais graves, a divulgação não autorizada de dados pode levar a danos à reputação pessoal. Além disso, a violação de privacidade pode também resultar em roubo de identidade, onde os dados pessoais são utilizados de forma fraudulenta para obter benefícios financeiros ou cometer crimes [Abdulhamid et al. 2014].

O problema da proteção da privacidade dos dados foi inicialmente proposto na década de 1970 [Dalenius 1977], que apontou que o objetivo de proteger informações privadas em uma base de dados é evitar que qualquer usuário, incluindo usuários legítimos e possíveis atacantes, obtenha informações precisas sobre indivíduos específicos ao acessar a base de dados. Com o passar dos anos, essa preocupação evoluiu significativamente, levando ao desenvolvimento de várias técnicas e metodologias para proteger a privacidade dos dados, como k-anonimato [Sweeney 2002], l-diversidade [Machanavajjhala et al. 2007], t-proximidade [Li et al. 2006], δ-presença [Nergiz et al. 2007], dentre outros. Contudo, cada um desses modelos oferece proteção apenas contra um tipo específico de ataque e não consegue se defender contra novos tipos de ataques que venham a ser desenvolvidos. A causa fundamental dessa deficiência reside no fato de que a efetividade de um modelo de preservação da privacidade depende da suposição de um conhecimento prévio específico por parte de um atacante, também denominado adversário. No entanto, é praticamente impossível enumerar todos os tipos possíveis de conhecimento prévio que um atacante pode obter. Portanto, é altamente desejável um modelo de preservação da privacidade que seja independente do conhecimento prévio do atacante.

A privacidade diferencial (PD) [Dwork 2006] surgiu como a noção padrão de privacidade, em vez de uma única ferramenta, para o compartilhamento de dados de maneira privada. Ela tem sido utilizada na indústria [Kenthapadi et al. 2019] por empresas como Apple, Google e Uber [Cormode et al. 2018], e também no setor público por agências dos EUA, como o U.S. Census Bureau [Garfinkel et al. 2018]. A privacidade diferencial assume que um atacante pode obter todas as informações de um conjunto de dados, exceto o registro alvo, o que pode ser considerado o conhecimento máximo que um atacante pode ter. Sob o conceito de privacidade diferencial, os resultados de análises realizadas em um conjunto de dados são insensíveis à alteração de um único registro, ou seja, a presença, ou ausência, de um único registro no conjunto de dados tem pouco efeito sobre a distribuição de saída das análises (consultas) realizadas. Em outras palavras, um atacante não pode obter informações precisas sobre um indivíduo ao observar os resultados das análises realizadas sobre um conjunto de dados diferencialmente privado, pois o risco de divulgação de privacidade causado pela adição, ou exclusão, de um único registro é mantido dentro de um intervalo aceitável especificado pelo usuário.

Existem diversos métodos e implementações para realizar análises de dados com privacidade diferencial. Embora tais abordagens tenham sido inicialmente projetadas para dados tabulares, a privacidade diferencial também pode ser aplicada à análise de dados em redes sociais [Silva et al. 2017, Jiang et al. 2021]. Em termos gerais, as redes sociais podem ser modeladas como grafos e se tornam extremamente complexas em larga escala. Dessa forma, existem desafios fundamentais que precisam ser enfrentados para aplicar a privacidade diferencial na análise de dados de redes sociais. Primeiro, é necessário estudar a privacidade diferencial de dados tabulares no contexto de dados de rede. Em seguida, é preciso abordar a questão da alta sensibilidade em dados de redes sociais complexas, uma vez que a presença de relações interdependentes entre elementos da rede pode amplificar o impacto das alterações de dados individuais, dificultando a proteção da privacidade. Por fim, é fundamental explorar o equilíbrio entre a utilidade dos dados para a análise e a garantia de privacidade, pois adicionar muito ruído para garantir a privacidade diferencial pode tornar os resultados das análises inúteis.

Este capítulo apresenta uma visão geral dos métodos e técnicas de privacidade diferencial destinados a proteger informações sensíveis, permitindo, simultaneamente, análises de dados relevantes em redes sociais. A Seção 4.2 apresenta os conceitos fundamentais sobre a análise de dados em redes sociais, enquanto a Seção 4.3 introduz problemas inerentes à privacidade dos indivíduos, decorrentes de análises indevidas sobre os dados. Em seguida, o modelo de privacidade diferencial é detalhado na Seção 4.4, com destaque para suas principais propriedades e configurações. Já a Seção 4.5 destaca as principais variantes do modelo de privacidade diferencial para o contexto de redes sociais, tais como *node*-PD, *edge*-PD, *edge-weight* PD e *attributed*-PD. Identificamos os principais tipos de análises realizadas sobre redes sociais, juntamente com as diversas técnicas de privacidade diferencial existentes para a realização dessas análises de maneira privada na Seção 4.6. Por fim, a Seção 4.7 explora as perspectivas futuras dos tópicos abordados neste capítulo, destacando as principais dificuldades a serem superadas.

4.2. Fundamentos de Análises de Dados em Redes Sociais

Na era digital, dados têm se transformado em ativos importantes para as organizações, principalmente devido ao seu elevado valor e importância. Atualmente, grandes volumes de dados, de diversas naturezas, encontram-se disponíveis e tornam-se grandes aliados estratégicos de empresas em seus processos de tomada de decisão a partir da análises realizadas. De maneira sucinta, a análise de dados consiste no processo de inspecionar, tratar, transformar e modelar os dados com o objetivo de descobrir informações úteis, *insights* e conclusões que auxiliem na tomada de decisão de empresas e organizações. A análise de dados dispõe de múltiplas técnicas e abordagens, abrangendo diferentes domínios [Mendonça et al. 2023]. A mineração de dados é uma das técnicas mais conhecidas utilizada para a análise de dados. Ela faz parte do grupo de análise preditiva, tendo como foco principal a modelagem estatística e a descoberta de conhecimento para fins preditivos. Juntamente com a análise preditiva, as análises prescritiva, descritiva e diagnóstica compõem os tipos de análise de dados existentes [Cook and Holder 2006].

Comumente, as análises de dados ocorrem sobre dados tabulares, representados por meio de registros, o que limita as análises sobre estruturas de dados mais complexas, como as redes sociais. Diferentemente das análises sobre dados tabulares, as análises sobre redes sociais priorizam os relacionamentos entre os indivíduos que compõem as redes e seus respectivos relacionamentos, ou conexões. Redes sociais são estruturas mais complexas que, intuitivamente, são modeladas a partir de estruturas em grafos. Em resumo, um grafo consiste em uma estrutura de dados formada, originalmente, por nós e arestas, onde os nós representam as entidades (indivíduos) e as arestas os relacionamentos entre nós. Vale ressaltar que neste capítulo, as análises de dados em redes sociais são conduzidas por meio de estruturas de grafos. Dessa forma, uma rede social é definida como um grafo $G = (V, E)$, onde V é o conjunto de vértices (nós) e E é o conjunto de arestas. Devido às características inerentes às redes sociais, diversas estatísticas podem ser extraídas de análises sobre grafos. As estatísticas mais comuns incluem os graus dos nós, juntamente com suas respectivas distribuição de graus, métricas de centralidade e outras medidas pertinentes. Adicionalmente, contagens de subgrafos, bem como diversas métricas de distância, também são exemplos de métricas frequentemente examinadas em análise de grafos.

4.2.1. Contagem de subgrafos

A análise de subgrafos desempenha um papel crucial na compreensão das redes sociais, proporcionando uma visão da estrutura e dos padrões subjacentes às interações entre os nós [Ribeiro et al. 2021]. Dentre os diferentes tipos de subgrafos, destacam-se os triângulos, estrelas e cliques, cada um fornecendo informações relevantes sobre a conectividade e a estrutura da rede. A Figura 4.1 exemplifica quatro tipos diferentes de subgrafos que podem estar presentes em uma rede social: triângulos, *k*-estrelas, *k*-cliques e *k*-triângulos.

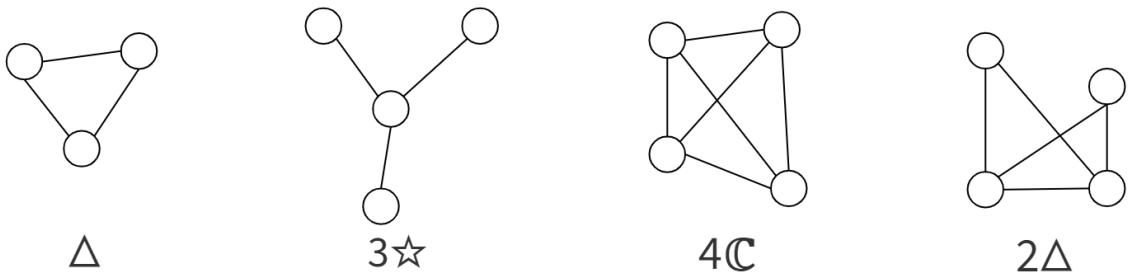


Figura 4.1: Exemplos de subgrafos que podem estar presentes em uma rede social: triângulos, estrelas de tamanho 3, cliques de tamanho 4 e triângulos de tamanho 2.

Os triângulos são subgrafos compostos por três nós interconectados, formando uma estrutura triangular. A contagem de triângulos em uma rede social é importante para identificar a presença de agrupamentos densamente conectados de três indivíduos. A ocorrência de triângulos indica relações de proximidade local, o que pode sugerir a existência de comunidades dentro da rede.

Por outro lado, uma *k*-estrela consiste em um nó central conectado a *k* nós periféricos, formando uma estrutura semelhante a uma estrela. Este conceito é particularmente valioso para identificar e caracterizar os nós mais influentes e os padrões de interação dentro de uma rede. Além disso, a distribuição de *k*-estrelas ajuda a compreender a dinâmica das redes sociais. Em redes de comunicação, por exemplo, um nó central com muitas conexões pode atuar como um *hub* de comunicação, onde a informação é agregada e distribuída.

Um *k*-clique é um subgrafo completo composto por *k* nós, onde cada par de nós está diretamente conectado por uma aresta. A contagem de *k*-cliques é essencial para identificar grupos coesos e altamente conectados na rede social, como grupos de amigos próximos ou equipes de trabalho colaborativo. A existência de *k*-cliques sugere que os membros desses subgrafos compartilham interesses comuns ou têm uma alta frequência de comunicação.

Os *k*-triângulos são uma extensão do conceito de triângulos em grafos. Em particular, esse subgrafo é formado por um conjunto de *k* triângulos que compartilham um vértice em comum. Esta estrutura é utilizada para identificar áreas de alta interconectividade em uma rede, onde múltiplas relações convergem em um único nó. A análise de *k*-triângulos é particularmente útil para identificar nós centrais ou altamente influentes na rede. Em redes de colaboração científica, por exemplo, um nó central com muitos *k*-triângulos pode representar um pesquisador que colabora com diversos grupos distintos, servindo como um ponto de interseção entre diferentes comunidades de pesquisa.

4.2.2. Histogramas

Os histogramas são representações gráficas que ilustram a distribuição de um conjunto de dados. Compostos por barras retangulares, cada barra representa a frequência de dados dentro de intervalos específicos, chamados *bins*. A altura das barras corresponde ao número de ocorrências dos dados em cada intervalo. Ao representar visualmente a distribuição de certas características de um grafo, os histogramas permitem uma compreensão clara de como a conectividade é estruturada dentro da rede. Eles ajudam a identificar padrões recorrentes e tendências que podem ser difíceis de perceber de outra maneira [Cook and Holder 2006]. Por exemplo, um histograma da distribuição de grau de um grafo revela como as conexões estão distribuídas entre os nós, destacando aqueles altamente conectados (*hubs*) e indicando se a rede segue uma distribuição de lei de potência ou se é mais homogênea. Um outro exemplo seria um histograma de pesos de arestas em grafos ponderados, no qual é possível identificar quais arestas carregam maior relevância ou influência na dinâmica da rede. Nesse cenário, arestas com pesos consistentemente altos, evidenciadas por picos no histograma, sugerem conexões críticas ou relacionamentos robustos entre os nós. A Figura 4.2 ilustra dois histogramas, um de grau dos nós e outro de peso das arestas, a partir de um grafo ponderado.

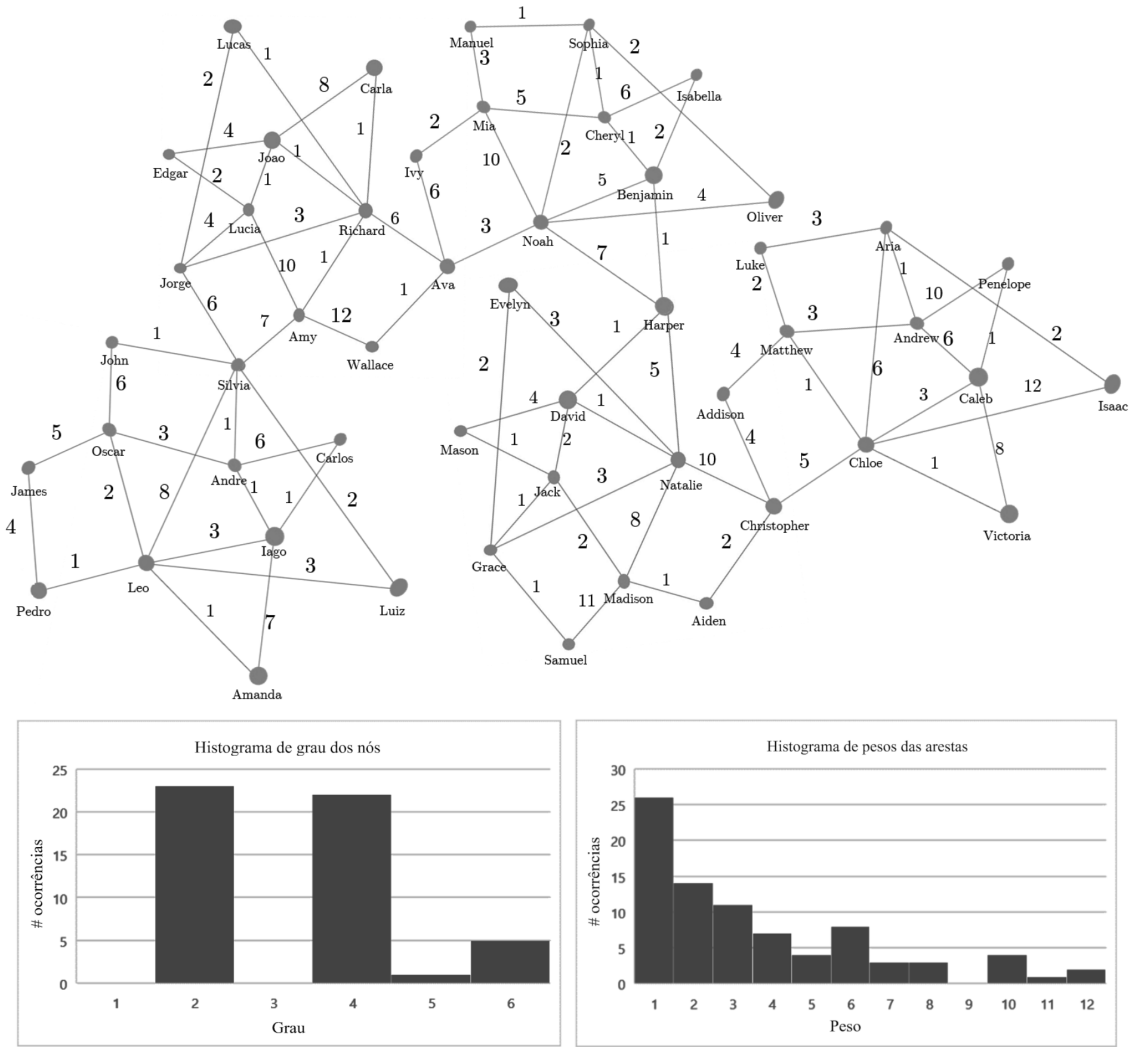


Figura 4.2: Histogramas de grau dos nós e de peso das arestas a partir de um dado grafo.

Adicionalmente, histogramas são úteis para examinar propriedades estruturais dos grafos, como a distribuição de coeficientes de agrupamento, que medem a tendência dos nós a formar *clusters*. Um histograma de coeficientes pode mostrar se a rede possui comunidades bem definidas ou se é mais dispersa. Outro aspecto crítico dos histogramas é sua capacidade de detectar anomalias. Ao comparar os histogramas de diferentes períodos ou subconjuntos da rede, é possível identificar mudanças súbitas ou padrões incomuns que possam indicar comportamentos anômalos, como ataques cibernéticos, falhas sistêmicas ou até mesmo a formação de novos grupos sociais inesperados. Por exemplo, uma alteração significativa na distribuição de grau pode sugerir a adição, ou remoção, massiva de nós ou arestas, apontando para eventos extraordinários na rede.

4.2.3. Centralidade

A centralidade é um conceito essencial na teoria de redes e análise de grafos, utilizado para medir a importância, ou influência, relativa de um nó dentro de uma rede [Bloch et al. 2023]. Essa métrica ajuda a identificar os nós mais centrais ou relevantes, que desempenham papéis cruciais na conectividade e dinâmica da rede. Existem várias medidas de centralidade, cada uma capturando diferentes aspectos da importância dos nós.

Centralidade de Grau (*Degree Centrality*): A centralidade de grau é a medida mais simples e é definida pelo número de conexões diretas (arestas) que um nó possui. Um nó com uma alta centralidade de grau é considerado popular ou altamente visível na rede, já que possui muitas conexões diretas com outros nós. Esta métrica é útil em redes sociais para identificar indivíduos com muitos contatos.

Centralidade de Proximidade (*Closeness Centrality*): A centralidade de proximidade mede a distância média de um nó a todos os outros nós na rede. Um nó com alta centralidade de proximidade pode alcançar todos os outros nós rapidamente, tornando-se central em termos de proximidade. Essa medida é particularmente útil para identificar nós que são eficientes na disseminação de informações pela rede.

Centralidade de Intermediação (*Betweenness Centrality*): A centralidade de intermediação quantifica o número de vezes que um nó atua como ponte ao longo dos caminhos mais curtos entre outros pares de nós. Nós com alta centralidade de intermediação são críticos para a comunicação dentro da rede, pois eles facilitam o fluxo de informações entre diferentes partes da rede.

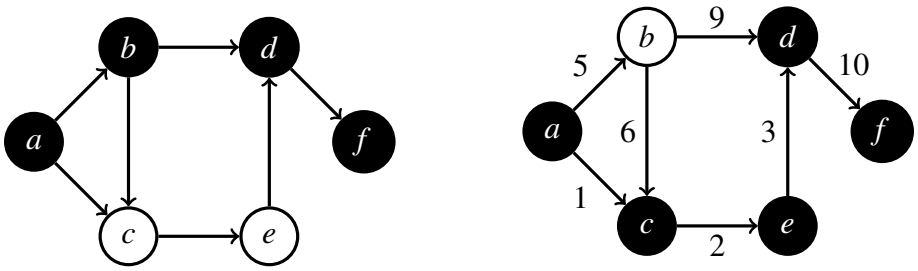
Centralidade de Autovetor (*Eigenvector Centrality*): A centralidade de autovetor mede a influência de um nó com base na importância dos seus vizinhos. Um nó com alta centralidade de autovetor está conectado a outros nós que também são altamente conectados. Essa métrica é útil para identificar não apenas nós centrais, mas também aqueles que estão ligados a outros nós influentes, proporcionando uma visão mais holística da influência dentro da rede.

Centralidade de PageRank (*PageRank Centrality*): Popularizada pelo algoritmo de ranqueamento (*ranking*) do Google, a centralidade de PageRank é semelhante à centralidade de autovetor, mas ajusta a influência de um nó pelo número de conexões que os seus vizinhos têm. É amplamente utilizada para determinar a importância de páginas *web* e pode ser aplicada a outras redes para identificar nós altamente influentes.

Cada uma das medidas de centralidade mencionadas proporciona uma perspectiva distinta sobre a estrutura e a dinâmica de uma rede, permitindo uma análise variada da importância dos nós. Ao combinar essas diferentes medidas, é possível obter uma compreensão ainda mais abrangente dos nós influentes e do papel que desempenham na rede. Dessa forma, a análise de centralidade é uma ferramenta poderosa para otimizar o desempenho, a robustez e a resiliência das redes em diversas aplicações, incluindo não somente redes sociais, mas também redes de comunicação, de transporte e financeiras.

4.2.4. Caminhos mínimos e distâncias

Um caminho mínimo entre dois nós em um grafo é definido como a sequência de arestas que conecta esses dois nós com a menor soma de arestas ou de pesos. Se o grafo é não ponderado, o caminho mínimo é simplesmente o caminho com o menor número de arestas. Já em grafos ponderados, o caminho mínimo é o que minimiza a soma dos pesos das arestas ao longo do caminho [Mitchell 2017]. A Figura 4.3 ilustra uma rede com caminho mínimo em um grafo não ponderado (Figura 4.3a) e em um grafo ponderado (Figura 4.3b).



(a) Grafo não ponderado. (b) Grafo ponderado.

Figura 4.3: Exemplo de caminhos mínimos entre os nós a e f.

Outra métrica importante envolvendo distâncias em grafos é a média de caminhos mínimos. Ela é calculada da seguinte forma: para cada caminho mínimo entre todos os pares de nós, são somadas todas as distâncias e esse valor é dividido pelo número total de pares de nós. Em redes sociais, uma média baixa de caminhos mínimos sugere uma rede mais coesa, onde os indivíduos estão mais próximos uns dos outros, facilitando a disseminação de informações, influências e interações. Por outro lado, uma média alta de caminhos mínimos indica uma rede mais dispersa, onde as distâncias entre os indivíduos são maiores. Tal fato pode dificultar a propagação rápida de informações.

Por fim, a métrica de diâmetro em um grafo é outra medida importante utilizada para caracterizar a distância máxima entre os vértices. Essa métrica mede a maior distância encontrada ao percorrer todos os pares de vértices no grafo, representando a distância mais longa que deve ser percorrida para conectar qualquer par de nós na rede. Diâmetros em redes sociais estão relacionados ao conceito de “seis graus de separação” [Samoylenko et al. 2023], o qual sugere que qualquer pessoa no mundo pode ser conectada a qualquer outra pessoa através de no máximo seis intermediários.

4.3. Ameaças de Privacidade em Redes Sociais

Ameaças de privacidade abrangem uma ampla gama de atividades que resultam no acesso a informações sensíveis por partes não autorizadas. Essas partes, que não deveriam ter acesso a tais informações, podem utilizá-las para fins mal-intencionados. Existem duas principais categorias de ataques de inferência observadas em redes sociais: a inferência de atributos privados [Mislove et al. 2010, Dey et al. 2012], que envolve a dedução de informações pessoais como gostos, interesses ou características demográficas; e a desanonimização (*de-anonymization)* [Narayanan and Shmatikov 2009, Brito et al. 2015, Qian et al. 2016], que visa identificar a identidade real de usuários anônimos, conectando dados anônimos a perfis pessoais conhecidos.

A inferência de atributos privados visa descobrir valores de atributos ocultos que são protegidos pelo usuário ou pelo curador dos dados. Os ataques de inferência baseados em vizinhança [Mislove et al. 2010, Dey et al. 2012, Gong et al. 2014] aproveitam o fato de que usuários próximos podem ter atributos semelhantes, ou idênticos, com alta probabilidade. Esses ataques inferem atributos privados de um usuário explorando os valores de atributos conhecidos de outros usuários com interesses similares [Chaabane et al. 2012]. Por exemplo, se a maioria dos amigos de um usuário trabalha como desenvolvedores de *software*, há uma grande chance de que o próprio usuário também trabalhe nessa área. Já a inferência baseada em comportamento tenta identificar semelhanças entre determinados valores de atributos através de dados comportamentais, como interesses, características pessoais e comportamentos culturais. Por exemplo, se a maioria dos filmes, séries e músicas que um usuário gosta são do gênero "suspense", é provável que o usuário tenha uma forte preferência por esse gênero.

A desanonimização [Narayanan and Shmatikov 2008, Narayanan and Shmatikov 2009, Qian et al. 2016] envolve o uso de uma rede social anonimizada e uma rede social de referência que contém as identidades verdadeiras dos indivíduos, mapeando os nós nesses dois grafos para que as identidades dos usuários no grafo anonimizado possam ser reidentificadas. Uma rede social anonimizada é geralmente disponibilizada por um curador dos dados a vários solicitantes, como pesquisadores, anunciantes, desenvolvedores de aplicativos e agências governamentais, após ocultar informações privadas identificáveis por meio de várias técnicas de anonimização. Uma rede social de referência pode ser obtida através de informações coletadas de outras fontes, como uma rede social diferente que compartilha usuários com a rede publicada. Geralmente, uma rede social de referência pode conter menos atributos sobre os nós do que uma rede social anonimizada, mas ainda assim pode ser usada para correlacionar e identificar indivíduos.

Essas duas categorias de ataques à privacidade resultam na exposição de diferentes tipos de informações sensíveis. A fim de proteger os dados privados em redes sociais e formalizar o conceito de privacidade nesse contexto, foram identificadas as principais ameaças à privacidade, conforme ilustrado na Figura 4.4. Essas ameaças são detalhadas a seguir.

Descoberta de identidade: Em redes sociais, a identidade de um indivíduo pode ser considerada privada, enquanto os atacantes podem explorar diversas informações para reidentificar um usuário da rede ou determinar se um indivíduo-alvo está presente, ou não, nela. Por exemplo, uma empresa disponibiliza para análise uma rede social profissional

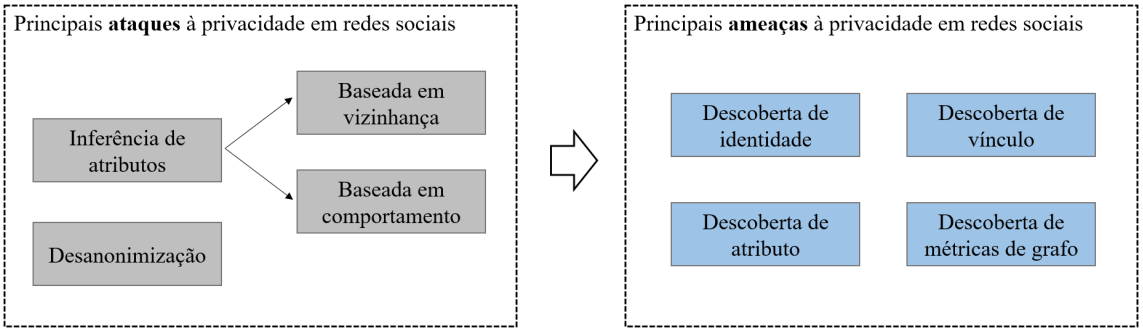


Figura 4.4: Principais ataques e ameaças à privacidade em redes sociais.

anonimizada, a qual contém Maria, com sua identidade oculta. Utilizando informações de outra rede social pública, um adversário observa que um colega de trabalho publica uma foto de uma festa de escritório em que Maria está presente, marcando-a pelo nome verdadeiro. Como resultado, a identidade de Maria pode ser inferida na rede social profissional anonimizada, comprometendo sua privacidade.

Descoberta de vínculo: As relações sociais entre indivíduos podem ser modeladas como arestas em um grafo. As informações de conexão podem ser consideradas sensíveis em alguns casos. Por exemplo, Kossinets e Watts [Kossinets and Watts 2006] analisaram um grafo derivado das comunicações por *e-mail* entre estudantes e membros do corpo docente em uma universidade, onde os relacionamentos de *e-mail* de "quem enviou *e-mails* para quem" foram considerados sensíveis. Um outro exemplo de descoberta de vínculo se dá no seguinte cenário: Ana é uma usuária ativa em uma rede social de *hobbies*, onde ela mantém um perfil anônimo para compartilhar seu interesse em pintura. Ela prefere manter sua associação com um grupo de discussão sobre arte em segredo de seu círculo social pessoal. No entanto, um algoritmo de recomendação de amigos da plataforma sugere conexões com seus colegas de trabalho, revelando indiretamente sua participação no grupo de discussão de arte e expondo seus interesses pessoais. Nesse caso, o algoritmo de recomendação compromete a privacidade de Ana ao expor sua participação no grupo de discussão de arte e seus interesses pessoais.

Descoberta de características do indivíduo: O perfil de um usuário em uma rede social comumente inclui uma variedade de atributos, como idade, gênero, área de estudo e ocupação. Alguns desses atributos, como salário, estado de saúde e informações sobre doenças, são considerados sensíveis e privados. Por exemplo, imagine uma plataforma de rede social voltada para saúde e bem-estar, onde os usuários compartilham informações sobre seus hábitos alimentares, rotinas de exercícios e condições médicas. João é um usuário ativo nessa plataforma, que compartilha regularmente detalhes sobre sua dieta, exercícios e algumas preocupações de saúde. Ele prefere manter sua condição médica de diabetes em segredo de seus colegas de trabalho e familiares, já que prefere tratar essa questão de forma privada. No entanto, João compartilha informações gerais sobre sua dieta saudável e rotina de exercícios na plataforma. A plataforma então decide lançar um recurso de análise de dados para os usuários, que fornece *insights* sobre hábitos saudáveis com base nas informações compartilhadas. Esse recurso utiliza algoritmos avançados para analisar os padrões de comportamento dos usuários na rede social e sugerir mudanças positivas

de estilo de vida. No entanto, durante a análise, o algoritmo da plataforma identifica João como um usuário com diabetes, com base em seus hábitos alimentares e padrões de atividade física. Consequentemente, a plataforma começa a fornecer sugestões específicas para gerenciar o diabetes, como dietas especiais e rotinas de exercícios, para todos os seus contatos na plataforma. Como resultado, João é confrontado com mensagens inesperadas de apoio e conselhos de seus colegas de trabalho e familiares, que ficaram sabendo de sua condição médica através das sugestões automáticas da plataforma.

Descoberta de métricas de grafo: Uma vez que as redes sociais podem ser representadas como grafos, métricas como grau, intermediação, centralidade, comprimento do caminho mais curto, contagem de subgrafos e peso de aresta podem ser utilizadas para realizar análises de redes sociais. No entanto, a divulgação dessas informações pode, inadvertidamente, resultar em vazamentos de dados pessoais. Por exemplo, considere uma rede social de uma comunidade universitária, onde os estudantes interagem entre si e com os professores por meio de postagens, comentários e mensagens privadas. A universidade está interessada em entender os padrões de interação entre os estudantes para melhorar a experiência educacional. A equipe da universidade decide usar métricas de grafo, como centralidade de grau, para identificar os alunos mais influentes na rede. Eles planejam usar essas informações para melhorar a distribuição de recursos, como atividades extracurriculares e oportunidades de liderança. Ao analisar as métricas de grafo, a equipe identifica Pedro como um dos alunos mais influentes na rede, com um alto grau de centralidade. Eles decidem então convidar Pedro para participar de um programa de mentoria para novos alunos, com base em sua influência percebida na comunidade. No entanto, Pedro não queria que sua posição na rede social fosse revelada dessa forma. Ele preferia manter sua participação discreta e demonstrou desagrado diante da atenção indesejada que recebeu como resultado da divulgação de suas métricas de grafo.

4.4. Introdução à Privacidade Diferencial

Nos últimos anos, a Privacidade Diferencial (PD) tornou-se o padrão para a preservação de privacidade em análises de dados sob fortes garantias matemáticas. Muito se deve às limitações existentes nos modelos de privacidade mais tradicionais, ou sintéticos, como o k-anonimato, l-diversidade, t-proximidade, δ-presença, dentre outros, os quais ainda deixavam os indivíduos vulneráveis quanto à sua privacidade [Brito and Machado 2017].

4.4.1. Intuição e Definição

Na definição original de privacidade diferencial, os dados privados são vistos como uma coleção de registros, onde cada registro corresponde a um indivíduo. Em essência, a privacidade diferencial assegura a proteção à privacidade do indivíduo ao injetar ruído no resultado de consultas aplicadas sobre os dados dos indivíduos, ou seja, modificando os dados originais através da introdução de aleatoriedade [Dwork et al. 2006]. A premissa base da privacidade diferencial é de que o resultado de qualquer consulta é, praticamente, igualmente possível de ocorrer, independente da presença, ou ausência, de qualquer indivíduo no conjunto de dados. É importante mencionar que a privacidade diferencial não é uma simples ferramenta, mas um paradigma capaz de quantificar e gerenciar os riscos de violação de privacidade. Portanto, a privacidade diferencial pode ser aplicada em desde simples estimativas estatísticas até mesmo em aprendizado de máquina.

Seja Q uma consulta a ser realizada sobre um conjunto de dados D, o qual contém informações sensíveis sobre um conjunto de indivíduos. A PD é definida através de um algoritmo aleatório M, também chamado de mecanismo, o qual é executado sobre D. A PD assegura que a saída de $M(D)$ deve ser semelhante à saída de $Q(D)$, ou seja, o objetivo da PD é fazer com que a saída de $M(D)$ seja o mais próximo possível à saída de $Q(D)$ para garantir a utilidade dos dados e, ao mesmo tempo, preservar a privacidade de todos os indivíduos presentes no conjunto de dados.

Para preservar a privacidade de todos os indivíduos através do mecanismo M, a PD estabelece a noção de conjuntos de dados vizinhos. Dois conjuntos de dados D e D' são ditos vizinhos se diferirem em, no máximo, um registro, denotado como $D \sim D'$. D' pode ser obtido a partir de D adicionando, ou removendo, um único registro. A PD também garante que, independentemente de quando a entrada seja D ou D', a probabilidade de uma determinada saída ocorrer a partir de $M(D)$ ou $M(D')$ é quase a mesma. Essa propriedade é denotada como indistinguibilidade de conjuntos de dados vizinhos. Em outras palavras, a PD afirma que qualquer resposta de uma consulta ocorre com probabilidade semelhante, independentemente da presença, ou ausência, de qualquer indivíduo no conjunto de dados.

Antes de apresentar a definição de PD, considere que $Range(M)$ consiste em todas as saídas possíveis de M, ou seja, seu domínio de saídas. Por exemplo, se M calcula o número de registros em um conjunto de dados, então $Range(M)$ é igual a um conjunto de números inteiros não negativos. Por fim, a definição de PD é apresentada abaixo:

Definição 1 (ε-Privacidade Diferencial [Dwork 2006])**.** *Um mecanismo M satisfaz ε-privacidade diferencial se, para quaisquer dois conjuntos de dados vizinhos D e D', e para qualquer saída possível $O \subseteq Range(M)$,*

$$\Pr[M(D) = O] \leq \exp(\varepsilon) \times \Pr[M(D') = O], \tag{1}$$

onde $\Pr[\cdot]$ *representa a probabilidade do mecanismo produzir a saída O.*

4.4.2. Orçamento de Privacidade

O parâmetro ε que aparece na Definição 1 é denominado orçamento de privacidade. Este parâmetro é responsável por controlar as diferenças entre as probabilidade das saídas de um mecanismo que é executado em dois conjuntos de dados vizinhos, ou seja, o orçamento de privacidade garante que essas diferenças sejam limitadas a no máximo ε.

O orçamento de privacidade consiste em um número real positivo que controla o nível de privacidade que um mecanismo M fornece. Um ε menor fornece garantias de privacidade mais fortes, com distribuições de probabilidade mais indistinguíveis, mas uma menor utilidade de dados, uma vez que mais ruído deve ser adicionado ao resultado. De maneira análoga, um ε maior fornece garantias de privacidade mais fracas, mas uma maior utilidade de dados.

Definir o valor adequado de ε para uma aplicação é uma tarefa bastante desafiadora. Tal tarefa exige o esforço de diversas partes, como especialistas em privacidade, *stakeholders* e proprietários de dados, ou seja, indivíduos que compartilham seus dados, a fim de fornecer *feedback* contínuo para garantir a privacidade dos indivíduos e, ao mesmo tempo, divulgar informações significativas. No entanto, o orçamento de privacidade normalmente assume valores pequenos, tornando as probabilidades de saída do mecanismo

quase as mesmas, independente da entrada do mecanismo ser D ou D'. Vários estudos já foram abordados com o objetivo de determinar um valor desejável para ε [Hsu et al. 2014, Li et al. 2016]. No entanto, tem sido bastante defendido que $0{,}1 \leq \varepsilon \leq 1$ fornece garantias de privacidade fortes e níveis de utilidade aceitáveis, enquanto $\varepsilon \geq 5$ é aceitável apenas em algumas aplicações específicas.

4.4.3. Sensibilidade

Conforme mencionado anteriormente, a PD pode ser alcançada ao adicionar uma quantidade apropriada de ruído aos resultados das consultas. Entretanto, adicionar ruído excessivo pode prejudicar drasticamente a utilidade dos dados, diminuindo a precisão das análises, enquanto uma quantidade insuficiente de ruído pode não fornecer as garantias de privacidade adequadas. Para isso, o ruído adicionado a uma consulta Q depende da sensibilidade global de Q [Dwork et al. 2006]. A definição de sensibilidade global é dada abaixo:

Definição 2 (Sensibilidade Global [Dwork et al. 2006])**.** *A sensibilidade global de uma consulta Q consiste na máxima distância l_1 entre as saídas de Q em quaisquer dois conjuntos de dados vizinhos D e D', dada por:*

$$\Delta Q = \max_{D,D'} ||Q(D) - Q(D')||_1. \tag{2}$$

A sensibilidade global, também chamada apenas de sensibilidade, mede o impacto máximo nos resultados da consulta a partir da adição, ou remoção, de qualquer registro no conjunto de dados. A sensibilidade serve como um parâmetro essencial para determinar a quantidade adequada de ruído adicionado à consulta. É importante mencionar que a sensibilidade está relacionada apenas à função de consulta e é independente do conjunto de dados. Por exemplo, uma consulta com baixa sensibilidade exige que apenas uma pequena quantidade de ruído seja adicionada aos resultados da consulta para mascarar o impacto da adição, ou remoção, de um registro. Por outro lado, quando a sensibilidade é alta, uma quantidade significativa de ruído deve ser adicionada aos resultados da consulta, a fim de garantir a privacidade dos indivíduos, comprometendo a utilidade dos dados.

Para algumas consultas, a sensibilidade é simples de ser calculada. Por exemplo, a sensibilidade de consultas de contagem é 1, uma vez que adicionar, ou remover, um registro no conjunto de dados afetará o resultado das consultas em no máximo 1. Por outro lado, a sensibilidade de consultas mais complexas, como consultas de máximo e soma, não é tão simples de calcular como as consultas de contagem.

Por exemplo, considere uma consulta que calcula a soma dos pesos das pessoas em um determinado conjunto de dados. A inclusão de um novo registro no conjunto de dados aumentará o resultado da consulta em um valor equivalente ao peso do indivíduo adicionado. Portanto, a sensibilidade dependerá do valor do peso do indivíduo adicionado ao conjunto de dados. Note que o mesmo raciocínio é valido para a remoção de um registro do conjunto de dados. Deseja-se, então, atribuir um valor específico para representar a sensibilidade dessa consulta, uma vez que a consulta deve ser independente do conjunto de dados. Para o domínio específico de pesos, existe um limite superior racional conhecido para o peso máximo que um indivíduo pode ter. De acordo com [Allardyce

2012], Jon Brower Minnoch foi a pessoa mais pesada conhecida no mundo já documentada, pesando impressionantes 635 quilos. Dessa forma, é plausível atribuir um valor de 635 à sensibilidade dessa consulta. No entanto, isto não serve como prova definitiva, uma vez que é impossível garantir que outra pessoa volte a pesar 635 quilos, ou mais. Então, em alguns domínios, determinar uma sensibilidade razoável pode ser uma tarefa desafiadora [Near and Abuah 2021].

4.4.4. Mecanismos

Mecanismos são algoritmos capazes de garantir as propriedades da PD. Para consultas numéricas, a PD pode ser alcançada através de vários mecanismos, como o mecanismo de Laplace [Dwork 2006] e o mecanismo geométrico [Ghosh et al. 2009]. Embora ambos os mecanismos tenham sido projetados para consultas numéricas, eles divergem no tipo de ruído que é adicionado ao resultado da consulta. O mecanismo de Laplace é recomendado para consultas que geram valores reais, pois esse mecanismo produz valores de ruído $\in \mathbb{R}$. Por sua vez, o mecanismo geométrico é recomendado para consultas que geram valores inteiros, pois esse mecanismo produz valores de ruído $\in \mathbb{Z}$. No entanto, nem todas as consultas geram valores numéricos. Para esse tipo de consulta, também chamada de consulta categórica, o mecanismo exponencial [McSherry and Talwar 2007] é mais adequado.

4.4.4.1. Mecanismo de Laplace

Conforme brevemente mencionado anteriormente, o mecanismo de Laplace adiciona ruído de valor real aos resultados da consulta. Como o nome sugere, o mecanismo depende da distribuição de Laplace para gerar valores aleatórios, os quais serão adicionados ao resultado da consulta. Seja x o ruído adicionado ao resultado de uma de consulta Q, a distribuição de Laplace é definida conforme:

Definição 3 (Distribuição de Laplace)**.** *A distribuição de Laplace com média 0 e escala b é a distribuição com função densidade de probabilidade*

$$Lap(x|b) = \frac{1}{2b} \cdot \exp(-\frac{|x|}{b}). \tag{3}$$

Considere $Lap(b)$ a distribuição de Laplace com escala b, Q uma consulta e D um conjunto de dados. O funcionamento do mecanismo de Laplace é dado por calcular o resultado de $Q(D)$ e perturbar esse resultado com a adição de um ruído gerado a partir da distribuição de Laplace. A escala b do ruído gerado é calibrada através da relação entra a sensibilidade da consulta e o orçamento de privacidade, de maneira que $b = \frac{\Delta Q}{\varepsilon}$.

Teorema 1 (Mecanismo de Laplace [Dwork et al. 2006])**.** *O mecanismo de Laplace que adiciona o ruído gerado a partir de* $Lap(\frac{\Delta Q}{\varepsilon})$ *satisfaz* ε*-PD.*

4.4.4.2. Mecanismo Geométrico

O mecanismo geométrico [Ghosh et al. 2009] consiste na versão discreta do mecanismo de Laplace, ou seja, adiciona ruído inteiro aos resultados das consultas seguindo a dis-

tribuição geométrica. Dessa forma, garante-se que o resultado final da consulta seja um número inteiro. Portanto, o mecanismo geométrico é especializado em melhorar o desempenho de consultas de contagem [Ghosh et al. 2009]. A distribuição geométrica é definida conforme:

Definição 4 (Distribuição Geométrica)**.** *Uma variável aleatória X gerada a partir da distribuição geométrica tem uma função massa de probabilidade*

$$P(X = x) = \frac{1-\alpha}{1+\alpha}\alpha^{|x|}, \tag{4}$$

onde $0 \leq \alpha \leq 1$.

Em particular, quando $\alpha = e^{-\frac{\varepsilon}{\Delta Q}}$, o mecanismo geométrico é ε-PD.

Teorema 2 (Mecanismo Geométrico [Ghosh et al. 2009])**.** *O mecanismo geométrico que adiciona o ruído gerado a partir da distribuição geométrica, com* $\alpha = e^{-\frac{\varepsilon}{\Delta Q}}$*, satisfaz* ε*-PD.*

4.4.4.3. Mecanismo Exponencial

Conforme mencionado anteriormente, o mecanismo exponencial surge como uma solução para consultas que não retornam valores numéricos. Para tanto, [McSherry and Talwar 2007] propuseram o mecanismo exponencial, o qual garante PD para consultas categóricas. Sua ideia principal consiste em escolher uma saída O do espaço de saída $\mathscr{O}$, de acordo com uma função de utilidade u. Essa função de utilidade atribui probabilidades de escolha exponencialmente maiores às saídas com utilidades mais elevadas. Além disso, a escolha de u depende da aplicação. Dessa forma, aplicações diferentes levam a funções de utilidade distintas. Diferente dos demais mecanismos anteriores, Laplace e geométrico, os quais geram seus ruídos proporcionalmente à sensibilidade da consulta Q, no mecanismo exponencial é utilizado o conceito de sensibilidade da função de utilidade para prover a saída do mecanismo. A sensibilidade da função de utilidade é definida abaixo:

Definição 5 (Sensibilidade Global da Função de Utilidade [McSherry and Talwar 2007])**.** *A sensibilidade global de uma função de utilidade u é dada por*

$$\Delta u = \max_{O \in \mathscr{O}} \max_{D,D' vizinhos} |u(D,O) - u(D',O)|. \tag{5}$$

Teorema 3 (Mecanismo Esponencial [McSherry and Talwar 2007])**.** *Dada uma função de utilidade* $u : (D \times \mathscr{O} \rightarrow \mathbb{Z})$ *para um conjunto de dados D, o mecanismo M que gera uma saída* $O \in \mathscr{O}$*, com probabilidade proporcional a* $\exp(\frac{\varepsilon \times u(D,O)}{2\Delta u})$*, satisfaz* ε*-PD.*

4.4.5. Propriedades

Várias propriedades úteis integram os mecanismos de PD, como o pós-processamento, a composição sequencial e a composição paralela. A propriedade de pós-processamento assume que qualquer função aplicada à saída de um mecanismo diferencialmente privado também satisfaz PD. Por outro lado, a propriedade de composição sequencial assume que

qualquer sequência de mecanismos diferencialmente privados, aplicados sobre o mesmo conjunto de dados, e que satisfaça PD isoladamente, também fornece PD. Por sua vez, a propriedade de composição paralela assume que um mesmo mecanismo diferencialmente privado, aplicado sobre conjuntos de dados disjuntos, e que satisfaça PD isoladamente, também fornece PD.

Teorema 4 (Pós-processamento [Dwork et al. 2014])**.** *Seja M qualquer mecanismo tal que $M(D)$ seja ε-diferencialmente privado, e seja f qualquer função. Então, $f(M(D))$ também satisfaz ε-PD.*

Teorema 5 (Composição Sequencial [Dwork et al. 2014])**.** *Considere que cada mecanismo M_i satisfaz ε_i-PD. A sequência de mecanismos diferencialmente privados $M_i(D)$ satisfaz $\sum \varepsilon_i$-PD.*

Teorema 6 (Composição Paralela [Dwork et al. 2014])**.** *Considere que cada conjunto de dados D_i é disjunto e que um mecanismo M satisfaz ε_i-PD para o conjunto de dados D_i. A sequência de mecanismos diferencialmente privados $M(D_i)$ satisfaz* $\max(\varepsilon_i)$*-PD.*

Quando combinadas, essas propriedades fornecem flexibilidade para desenvolver uma maneira de agregar várias etapas diferencialmente privadas em um único mecanismo que satisfaça a PD.

4.4.6. Privacidade Diferencial Local

A privacidade diferencial, conforme apresentada anteriormente, considera a existência de um curador confiável (*third-party*), o qual é responsável por coletar os dados, perturbar os resultados das consultas através de um mecanismo que satisfaça a PD e disponibilizar os resultados ruidosos. Essa configuração de PD é geralmente denominada de PD global, PD centralizada, ou simplesmente PD. No entanto, encontrar um curador confiável para coletar e processar os dados pode ser uma tarefa bastante desafiadora em cenários práticos. Portanto, a falta de curadores confiáveis restringe a aplicabilidade da PD global. Por conta disso, a Privacidade Diferencial Local (PDL) [Duchi et al. 2013] foi proposta como uma abordagem diferencialmente privada que desconsidera a necessidade de um curador de dados confiável. Dessa forma, em vez de centralizar o fluxo de dados em uma única entidade externa supostamente confiável, cada indivíduo é responsável por proteger os seus próprios dados, perturbando-os localmente, por meio de um mecanismo diferencialmente privado, antes de enviá-los ao curador de dados. Na PDL, o curador de dados também é comumente chamado de *agregador*.

Quando comparada à PD global, a PDL é uma noção mais forte de privacidade, a qual mantém os dados sensíveis dos indivíduos privados até mesmo de curadores de dados não confiáveis. No entanto, por se tratar de uma noção de privacidade mais forte, espera-se que a PDL introduza mais ruído aos resultados sob as mesmas circunstâncias, ou seja, utilizando-se o mesmo orçamento de privacidade, em comparação com o PD global. A definição formal da PDL é formalmente apresentada abaixo:

Definição 6 (ε-Privacidade Diferencial Local)**.** *Um mecanismo M satisfaz ε-privacidade diferencial local se, para qualquer par de valores de entrada v e v', e para qualquer saída possível $O \subseteq Range(M)$:*

$$\Pr[M(v) = O] \leq \exp(\varepsilon) \times \Pr[M(v') = O]. \tag{6}$$

A principal diferença entre a PD e a PDL está nos dados de entrada que os mecanismos recebem. Um mecanismo de PD global recebe um conjunto de dados D como entrada, ou seja, os dados de todos os indivíduos, e garante que a saída seja indistinguível, enquanto que a PDL recebe apenas os dados de um único indivíduo v como entrada e gera respostas ruidosas por indivíduo, de maneira independente.

4.4.7. Protocolos de Privacidade Diferencial Local

Como mencionado anteriormente, os mecanismos são formas de garantir as propriedades de PD. Na PDL, essas propriedades são alcançadas através do uso de protocolos. Portanto, na PDL, os mecanismos são chamados de protocolos. Em resumo, protocolos são técnicas que modificam os dados do indivíduo para garantir as propriedades da PDL. O fluxo padrão de um protocolo de PDL consiste em: (I) codificar os dados do indivíduo; (II) perturbar os dados do indivíduo; e (III) enviar os dados ruidosos do indivíduo ao curador de dados.

(I) Codificação: Nessa etapa, os dados v do indivíduo são codificados em um vetor de bits B de tamanho d formado por 0's e 1's, de maneira que o valor 1 é atribuído às posições do vetor B que correspondem a v e 0 para as demais. Portanto, define-se a função `Codificar`(v) = B, tal que $B[v] = 1$ e $B[i] = 0$ para todo $i \neq v$.

(II) Perturbação: Nessa etapa, o vetor de bits codificado B é perturbado de acordo com dois parâmetros principais: p e q, formando um novo vetor de bits B', conforme mostrado na Equação 7. O parâmetro p consiste na probabilidade de que um bit i de B atribuído com o valor 1 permaneça sendo 1 mesmo após ser perturbado, ou seja, $B[i] = 1 \rightarrow B'[i] = 1$. Por sua vez, q é a probabilidade de que um bit de B atribuído com 0 se torne 1 após ser perturbado, ou seja, $B[i] = 0 \rightarrow B'[i] = 1$. Intuitivamente, $(1-p)$ e $(1-q)$ representam as probabilidade de $B[i] = 1 \rightarrow B'[i] = 0$ e $B[i] = 0 \rightarrow B'[i] = 0$, respectivamente. Essa etapa é executada de maneira diferencialmente privada, através do uso do parâmetro de orçamento de privacidade ε para determinar os valores de probabilidade p e q. Além disso, os valores de p e q dependem não apenas do valor de ε, mas também do protocolo escolhido. Uma vez perturbado, o vetor B' é reportado ao agregador.

$$\Pr[B'(i) = 1] = \begin{cases} p, \text{ se } B[i] = 1 \\ q, \text{ se } B[i] = 0 \end{cases} \tag{7}$$

(III) Agregação: Nessa etapa, o agregador coleta todos os vetores de bits perturbados B' reportados pelos indivíduos, e realiza a análise desses dados a partir de informações agregadas. A base para realizar as análises consiste em identificar o número de ocorrências de cada possível valor de entrada v a partir dos vetores B', através de uma função de `Suporte`. Por exemplo, um vetor B' é dito que suporta um valor de entrada v se $B'[v] = 1$, ou seja, `Suporte`$(B') = \{v \mid B'[v] = 1\}$ é conjunto de valores presentes em B'. De maneira semelhante, `Suporte`(v) é definido como o número de ocorrências do valor v nos vetores B' reportados.

É importante mencionar que as funções de `Codificar` e `Suporte` são diretamente dependentes do protocolo de PDL utilizado. Dessa forma, protocolos diferentes podem implementar essas funções de maneira diferente. Além disso, algumas informa-

ções relevantes são de domínio público, ou seja, conhecidas pelo agregador. Em resumo, o número de respostas n, o tamanho do vetor codificado d, o orçamento de privacidade ε e o protocolo de PDL utilizado, todas essas informações são de conhecimento do agregador. Dessa forma, o agregador é capaz de compreender a partir de qual protocolo de PDL que os dados foram sanitizados e, então, calcular os respectivos valores de probabilidade p e q. Finalmente, o agregador pode realizar uma estimativa imparcial da frequência dos valores reportados de acordo com o Teorema 7. Quando cada indivíduo envia os seus dados apenas uma única vez, o número de respostas n pode ser tratado como o número de indivíduos de maneira equivalente.

Teorema 7 (Estimativa Imparcial [Wang et al. 2017])**.** *Dado um protocolo de PDL, o número de ocorrências (contagem) de um valor v, dado por* $\tilde{c}(v) = \frac{\texttt{Suporte}(v) - n \times q}{p - q}$*, é imparcial, onde* $\texttt{Suporte}(v)$ *é o número de respostas que contém o valor v e n é o número de respostas.*

O problema de estimativa de frequências, onde o agregador busca estimar as frequências dos valores em um domínio previamente estabelecido, é um dos problemas mais fundamentais que a PDL se propõe a resolver. Problemas dessa natureza são comumente conhecidos como Oráculos de Frequência (OF). Vários estudos já foram realizados para desenvolver protocolos de OF [da Costa Filho and Machado 2023, Acharya et al. 2019, Bassily and Smith 2015, Ye and Barg 2018], onde o Protocolo de Resposta Aleatória (PRA) [Dwork et al. 2006] e o Protocolo de Codificação Unária (PCU) [Erlingsson et al. 2014] estão entre os mais disseminados na literatura.

4.4.7.1. Protocolo de Resposta Aleatória (PRA)

O protocolo de resposta aleatória foi um dos primeiros protocolos de OF propostos na literatura. Dentre as suas principais características, destaca-se a de permitir que um valor v seja codificado em um vetor de bits B, de maneira que B possua mais de um bit representativo, ou seja, marcado com 1. Tal representação pode ser bastante útil em diversos domínios. Imagine, dentro do cenário de redes sociais, que um indivíduo u deseja reportar as suas conexões existentes com outros indivíduos. Nesse caso, o valor v a ser reportado consiste em uma lista contendo os demais indivíduos que se conectam com u. Portanto, uma maneira de codificar v seria transformá-lo em em um vetor de bits B, de modo que $B[i] = 1$ indica que existe uma conexão entre os indivíduos u e i, enquanto $B[i] = 0$ indica a inexistência dessa conexão.

Visando garantir as propriedades da PDL, foi provado em [Erlingsson et al. 2014] que, para que o PRA satisfaça ε-PDL, a etapa de perturbação precisa ser realizada com valores específicos de p e q, tal que $p = \frac{1}{1+e^{\varepsilon}}$ e $q = 1 - p$.

4.4.7.2. Protocolo de Codificação Unária (PCU)

O protocolo de codificação unária difere-se do protocolo de resposta aleatória, principalmente, na forma em como é realizada a codificação dos dados. Como o nome sugere, no PCU, a codificação de qualquer valor é realizada através de um único bit representativo.

Assim, para um dado valor v, este será codificado em um vetor de bits B, de maneira que B possua um único bit marcado com 1, enquanto todos os demais bits assumem um valor igual a 0. Tal representação também é bastante útil em diversos domínios, principalmente em domínios mais simples e de menor dimensão d.

Assim como no PRA, também é necessário estabelecer quais são os valores de p e q que permitem que o PCU satisfaça ε-PDL, afim de garantir as propriedades da PDL. Diante disso, surgiram alguns protocolos baseados no protocolo de codificação unária com características particulares. Dentre eles, destacam-se o Protocolo de Codificação Unária Simétrica (PCUS) [Erlingsson et al. 2014] e o Protocolo de Codificação Unária Otimizada (PCUO) [Wang et al. 2017]. Ambos os protocolos são bastante similares em relação as etapas executadas por cada um, a grande diferença está na escolha dos parâmetros p e q a serem utilizados na etapa de perturbação de cada protocolo.

No protocolo de PCUS, os valores de p e q assumem valores que tratam os bits iguais a 0 e 1 de maneira simétrica. Os valores de p e q são dados por $p = \frac{e^{\frac{\varepsilon}{2}}}{e^{\frac{\varepsilon}{2}}+1}$ e $q = \frac{1}{e^{\frac{\varepsilon}{2}}+1}$, de maneira que $p+q = 1$. Por sua vez, o PCUO consiste em um melhoramento do PCUS, o qual propõe valores ótimos para p e q. Dessa forma, atribuir $p = \frac{1}{2}$ e $q = \frac{1}{e^{\varepsilon}+1}$ potencializa a utilidade das frequências estimadas. Note que, o valor de $p+q$ nunca será igual a 1, no máximo será próximo, visto que o valor de ε é sempre positivo. Em resumo, a ideia do PCUO é de que, com esses valores de p e q, o protocolo tenta intensificar que os bits reportados como 1 sejam, de fato, aqueles que originalmente eram 1, assim como os bits iguais a 0 são aqueles que, originalmente, também eram iguais a 0.

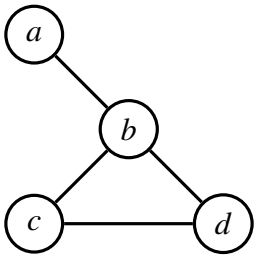
Por fim, é importante destacar que não existe um protocolo que seja o melhor para tudo. Existem diversos protocolos na literatura, com características e finalidades diferentes. Portanto, determinar qual protocolo é mais recomendado para um determinada tarefa se torna uma tarefa bastante desafiadora [Wang et al. 2017].

4.5. Privacidade Diferencial para Redes Sociais

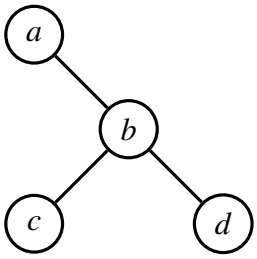
O conceito fundamental de privacidade diferencial depende diretamente da definição de conjuntos de dados vizinhos. Nas definições anteriores, um conjunto de dados vizinho é definido como um conjunto de dados obtido pela adição, ou remoção, de um único registro. Entretanto, no contexto de redes sociais, que concentram-se principalmente nas relações entre indivíduos, a associação entre dados privados e os registros dos conjuntos de dados tornam-se menos aparente. Dessa forma, antes de podermos aplicar PD em redes sociais, é necessário estabelecer uma nova definição de conjuntos de dados vizinhos que considere a estrutura de rede social e a semântica de privacidade associada à rede.

4.5.1. *Edge*-Privacidade Diferencial

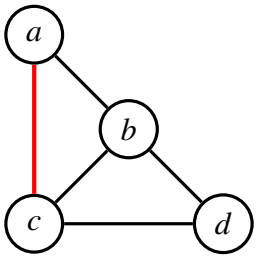
A primeira configuração de PD para redes sociais é chamada de *edge*-privacidade diferencial (*edge*-PD) [Hay et al. 2009]. A noção de *edge*-PD considera que duas redes sociais são vizinhas se uma puder ser obtida a partir da outra ao adicionar, ou remover, uma única aresta, ou adicionar, ou remover, um único usuário isolado, ou seja, um nó sem nenhuma aresta conectada a ele. A definição de *edge*-PD é formalmente definida abaixo:



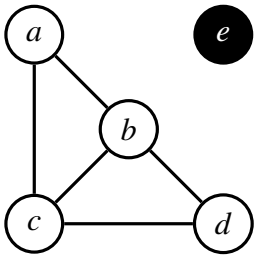
(a) Rede social G.



(b) Rede social vizinha G'_1 obtida a partir de G após a remoção da aresta (c,d).



(c) Rede social vizinha G'_2 obtida a partir de G após a adição da aresta (a,c).



(d) Rede social vizinha G'_3 obtida a partir de G após a adição do usuário e sem nenhuma aresta.

Figura 4.5: Exemplo de uma rede social e suas possíveis redes sociais vizinhas de acordo com a noção *edge*-PD.

Definição 7 (ε-*Edge*-Privacidade Diferencial [Hay et al. 2009])**.** *Um mecanismo M satisfaz ε-edge privacidade diferencial se, para qualquer par de grafos $G = (V,E)$ e $G' = (V',E')$, tal que $|V \oplus V'| + |E \oplus E'| = 1$, e para qualquer saída possível $O \subseteq Range(M)$,*

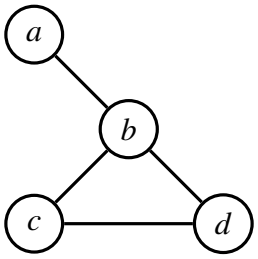
$$\Pr[M(G) = O] \leq \exp(\varepsilon) \times \Pr[M(G') = O]. \tag{8}$$

Dessa forma, um algoritmo que garante *edge*-PD fornece proteção contra a descoberta de arestas dos usuários, ou seja, os relacionamentos. Portanto, é importante destacar que esse nível de privacidade pode ser adequado para algumas aplicações. No entanto, existem alguns cenários em que é desejável estender as garantias de privacidade para além dos relacionamentos dos usuários.

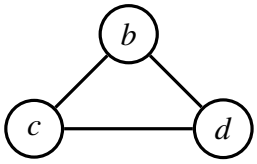
A Figura 4.5 apresenta um exemplo de uma rede social G e três possíveis redes sociais vizinhas G'_1, G'_2 e G'_3, de acordo com a noção de vizinhança provida na *edge*-PD. A rede G é composta, inicialmente, por 4 nós, ou usuários, e 4 arestas representando os relacionamentos entre os usuários (Figura 4.5a). Dentre as possíveis redes sociais vizinhas, a rede G'_1 é gerada a partir da adição de uma aresta (Figura 4.5b), enquanto a rede G'_2 é gerada a partir da remoção de uma aresta (Figura 4.5c) e, por fim, a rede G'_3 é gerada a partir da adição de um nó sem nenhuma uma aresta (Figura 4.5d).

4.5.2. *Node*-Privacidade Diferencial

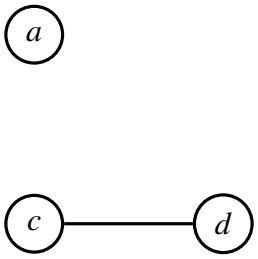
Proposto por [Kasiviswanathan et al. 2013], a *node*-privacidade diferencial (*node*-PD) consiste em uma noção mais estrita de PD quando comparada à *edge*-PD, visto que obje-



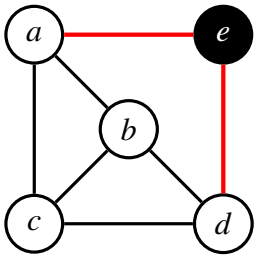
(a) Rede social G.



(b) Rede social vizinha G'_1 obtida a partir de G após a remoção do usuário a, juntamente com todas as suas arestas adjacentes (a,b).



(c) Rede social vizinha G'_2 obtida a partir de G após a remoção do usuário b, juntamente com todas as suas arestas adjacentes (b,a), (b,c) e (b,d).



(d) Rede social vizinha G'_3 obtida a partir de G após a adição do usuário e, juntamente com todas as suas arestas adjacentes (e,a) e (e,d).

Figura 4.6: Exemplo de uma rede social e suas possíveis redes sociais vizinhas de acordo com a noção *node*-PD.

tiva limitar a inferência sobre a existência, ou ausência, de um usuário em uma rede social. Assim, a *node*-PD fornece não somente garantias de privacidade aos usuários, mas também a todos os seus relacionamentos adjacentes. Assim, duas redes sociais G e G' são consideradas vizinhas se diferirem em um único nó e todas as suas arestas adjacentes. A definição de *node*-PD é formalmente apresentada abaixo:

Definição 8 (ε-*Node*-Privacidade Diferencial [Kasiviswanathan et al. 2013])**.** *Um mecanismo M satisfaz ε-node privacidade diferencial se, para qualquer par de grafos $G = (V,E)$ e $G' = (V',E')$, tal que $|V \oplus V'| = 1$ e $E \oplus E' = \{(u,v) | u \in (V \oplus V') \text{ ou } v \in (V \oplus V')\}$, e para qualquer saída possível $O \subseteq Range(M)$,*

$$\Pr[M(G) = O] \leq \exp(\varepsilon) \times \Pr[M(G') = O]. \tag{9}$$

Alcançar a PD no modelo de privacidade *node*-PD é muito mais difícil do que na *edge*-PD, uma vez que a *node*-PD fornece garantias de privacidade mais fortes. Dessa forma, pode ser que seja inviável projetar algoritmos que garantam a *node*-PD e, simultaneamente, forneçam análises precisas em redes sociais. A Figura 4.6 apresenta um exemplo de uma rede social G e três possíveis redes sociais vizinhas G'_1, G'_2 e G'_3, de acordo com a noção de vizinhança provida na *node*-PD.

A rede G é composta, inicialmente, por 4 nós, ou usuários, e 4 arestas representando os relacionamentos entre os usuários (Figura 4.6a). Dentre as possíveis redes

sociais vizinhas, as redes G_1' e G_2' são geradas a partir da remoção de um nó, juntamente com todas as suas arestas adjacentes (Figuras 4.6b e 4.6c), enquanto a rede G_3' é gerada a partir da adição de um nó, juntamente com todas as suas arestas adjacentes (Figura 4.6d).

4.5.3. *Edge-weight* Privacidade Diferencial

As duas principais alternativas propostas para aplicar privacidade diferencial em grafos, *node*-privacidade diferencial e *edge*-privacidade diferencial, não são adequadas quando os grafos são ponderados. Em geral, não é possível disponibilizar algumas informações como, por exemplo, caminhos mínimos, com um nível de utilidade significativo através das noções de *node*-PD ou *edge*-PD, uma vez que modificar uma única aresta pode alterar as distâncias do grafo significativamente. Dessa forma, surge um modelo de privacidade diferencial adequado para grafos ponderados, o qual é denominado *edge-weight* privacidade diferencial (*edge-weight* PD).

Nesse contexto, existem dois tipos principais de *edge-weight* PD, um que considera que a topologia de um grafo é conhecida [Sealfon 2016] e outro que a considera desconhecida [Brito et al. 2023]. Suponha $G = (V, E, \omega)$ um grafo ponderado não direcionado com uma função de peso $\omega : V^2 \to \mathbb{R}^+$ que mapeia conexões entre um par de vértices (u, v) para pesos em G.

Definição 9 (Funções de peso vizinhas com topologia conhecida [Sealfon 2016])**.** *Duas funções de peso ω, $\omega' : V^2 \to \mathbb{R}^+$ são vizinhas, denotadas por $\omega \sim \omega'$, se:*

$$||\omega - \omega'||1 := \sum u, v \in V |\omega(u, v) - \omega'(u, v)| \leq 1. \tag{10}$$

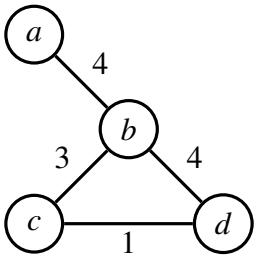
Dois grafos ponderados G e G' são vizinhos se possuem o mesmo conjunto de vértices e arestas, e se suas funções de peso diferem em uma unidade.

Definição 10 (Grafos ponderados vizinhos com topologia conhecida [Sealfon 2016])**.** *Considere $G = (V, E, \omega)$ e $G' = (V', E', \omega')$ dois grafos ponderados. G e G' são vizinhos se $V = V'$, $E = E'$ e $\omega \sim \omega'$.*

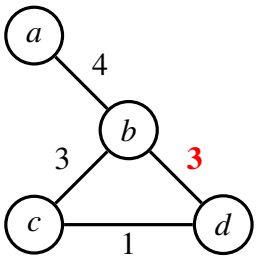
Vários trabalhos na literatura utilizam essas definições em seus estudos [Li et al. 2017, Pinot et al. 2018, Wang and Long 2019, Chen et al. 2022, Fan and Li 2022].

Para algumas aplicações do mundo real, a suposição de que a topologia do grafo é pública pode não ser verdadeira. Por exemplo, ao proteger a presença, ou ausência, de interações em uma rede de contatos entre dispositivos *IoT*, ou a existência, ou ausência, de chamadas telefônicas, mensagens de texto, ou a presença, ou ausência, de coautoria em um artigo. Esses tipos de interações não são cobertos pelas definições de Sealfon em termos de privacidade. Uma vez que uma aresta já é conhecida, qualquer mecanismo diferencialmente privado não mudará a presença, ou ausência, dessa conexão. Dessa forma, considerar apenas o cenário onde a topologia do grafo é publicamente conhecida não é eficaz para fornecer as garantias de privacidade desejadas.

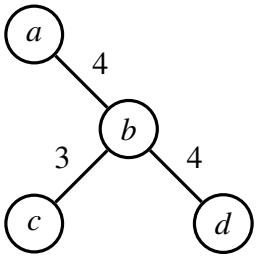
Para abordar essa limitação, o trabalho em [Brito et al. 2023] adapta a noção proposta por Sealfon para funções de peso vizinhas e fornece uma nova definição para grafos ponderados vizinhos com topologia desconhecida. Dessa forma, Seja $G = (V, E, \omega)$ um



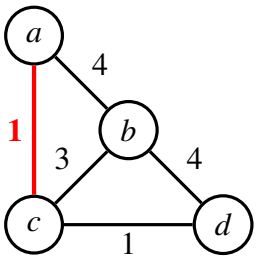
(a) Rede social ponderada G.



(b) Rede social vizinha G'_1 obtida a partir de G após a subtração de uma unidade de peso da aresta (b,d).



(c) Rede social vizinha G'_2 obtida a partir de G após a remoção da aresta (c,d) de peso igual a 1.



(d) Rede social vizinha G'_3 obtida a partir de G após a adição da aresta (a,c) de peso igual a 1.

Figura 4.7: Exemplo de uma rede social e suas possíveis redes sociais vizinhas de acordo com a noção *edge-weight* PD com topologia desconhecida.

grafo não direcionado e ponderado e uma função de peso $\omega : V^2 \to \mathbb{Z}_{\geq 0}$ que mapeia conexões (interações) entre um par de vértices (u,v) para pesos em G. O par $(u,v) \in E$ se os vértices u e v compartilham uma aresta comum, e $(u,v) \notin E$, caso contrário. Se $(u,v) \notin E$, então $\omega(u,v) = 0$. Como G é não direcionado, $\omega(u,v) = \omega(v,u)$.

Definição 11 (Funções de peso vizinhas com topologia desconhecida [Brito 2023])**.** *Duas funções de peso ω, $\omega' : V^2 \to \mathbb{Z}_{\geq 0}$ são vizinhas, denotado por $\omega \sim \omega'$, se:*

$$||\omega - \omega'||1 := \sum u, v \in V |\omega(u,v) - \omega'(u,v)| = 1. \tag{11}$$

A Definição 11 difere da Definição 9 no sentido de que agora os pesos podem assumir valores zero e os pesos também são valores inteiros. Assim, dois grafos G e G' são considerados vizinhos se tiverem o mesmo conjunto de vértices e se as funções de peso diferirem em uma unidade, conforme definição a seguir:

Definição 12 (Grafos ponderados vizinhos com topologia desconhecida [Brito 2023])**.** *Sejam $G = (V,E,\omega)$ e $G' = (V',E',\omega')$ dois grafos ponderados, G e G' são vizinhos se $V = V'$ e $\omega \sim \omega'$.*

A Figura 4.7 mostra um exemplo de três grafos ponderados vizinhos com topologia desconhecida a partir de uma rede G de entrada.

Com base nas definições apresentadas anteriormente, a definição formal de privacidade diferencial para grafos ponderados, considerando arestas e pesos como sendo informações privadas, é descrita como:

Definição 13 (ε-*Edge-weight* Privacidade Diferencial [Brito 2023])**.** *Um mecanismo M satisfaz ε-privacidade diferencial de pesos das arestas, se para qualquer par de grafos $G = (V, E, \omega)$ e $G' = (V', E', \omega')$, tais que G e G' são grafos de pesos vizinhos e para qualquer possível saída $O \subseteq Range(M)$,*

$$Pr[M(G) = O] \leq \exp(\varepsilon) \times Pr[M(G') = O]. \tag{12}$$

4.5.4. *Attributed*-Privacidade Diferencial

Em cenários reais, as redes sociais raramente são representadas apenas por nós e seus relacionamentos diretos. Em sua grande maioria, as redes sociais são representadas por uma complexa estrutura de dados, a qual consiste em uma fonte de informação extremamente rica. É bastante comum observarmos informações adicionais associadas à estrutura do grafo, sejam nos nós, ou nas arestas. A existência dessas informações, seja nas arestas, ou nos nós, consiste em informações valiosas para a compreensão e explicação de diversos comportamentos dos usuários. Assim, denominam-se esses grafos de grafos com atributos.

Grafos com atributos são uma classe particular de grafos nas quais informações adicionais, também chamadas de atributos, são anexadas à estrutura do grafo. Como antecipado anteriormente, os atributos podem estar associados tanto às arestas, quanto aos nós do grafo. Em relação à natureza dos atributos, estes possuem uma grande flexibilidade, podendo ser de qualquer natureza, seja numérica, categórica, lógica, dentre outras. Além disso, é importante destacar que os atributos são se limitam à uma única informação, de maneira que mais de um atributo pode estar relacionado às arestas, ou nós, do grafo.

Nesse contexto, os grafos passam a ser definido por $G = (V, E, X)$, onde V e E continuam sendo os conjuntos de vértices (nós) e arestas, respectivamente. Já o novo conjunto, denotado por X, consiste no conjunto de atributos associados à estrutura do grafo. Em um grafo com atributos nas arestas, X representará o conjunto dos atributos associados a cada aresta de E. Analogamente, em um grafo com atributos nos nós, X representará o conjunto dos atributos associados a cada nó de V.

Os grafos com atributos nas arestas têm sido amplamente adotados em diversos campos para explicar os motivos que levam os usuários a possuírem conexões entre si. Em redes sociais dessa natureza, o tipo de relacionamento entre os usuários da rede é de extrema importância. Alguns exemplos incluem redes de comunicação [Wang et al. 2013], redes de coautoria [Alsmadi and Alhami 2015] e redes de informação heterogênea [Shi et al. 2016]. Assim, o estudo de grafos com atributos nas arestas transformou-se em uma área de pesquisa próspera, tornando-se relevante para diversas aplicações como: detecção de anomalias [Shah et al. 2016], análise de mobilidade [Kaytoue et al. 2017] e pesquisa de comunidade [Li et al. 2023].

A Figura 4.8 apresenta um exemplo de um grafo com atributos nas arestas, onde os nós representam os colaboradores de uma empresa e as arestas afirmam a existência de

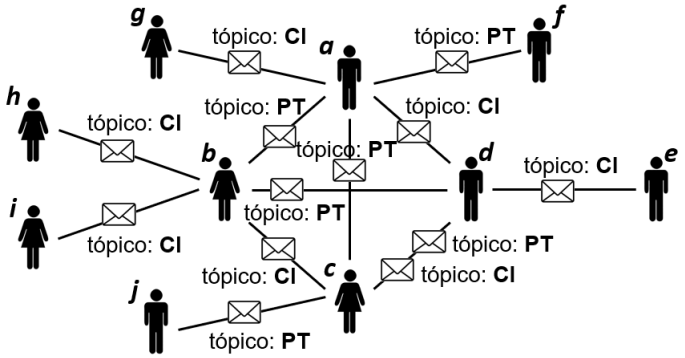


Figura 4.8: Grafo com atributos nas arestas, onde os nós representam os usuários e as arestas os *e-mails* trocados entre os usuários. O tópico dos *e-mails* trocados representa o atributo da aresta, onde "CI" denota assuntos relacionados a Comunicação Interna e "PT" retrata assuntos relacionados a Problemas Técnicos.

e-mails trocados entre dois colaboradores, juntamente com o tópico do *e-mail*. Os tópicos dos *e-mails* limitam-se a Comunicação Interna (CI) e Problemas Técnicos (PT), sendo detalhados a seguir.

- **Comunicação Interna (CI):** Utilizado para a comunicação interna da equipe, incluindo atualizações de projetos, reuniões e informações em geral.
- **Problemas Técnicos (PT):** Utilizado para questões relacionadas a problemas técnicas, onde problemas com plataformas de serviços e utilização de *softwares* são os mais comuns.

Por sua vez, redes sociais com atributos nos nós são bastante estudadas em tarefas de predição, detecção de padrões, descoberta de subgrafos e nós discrepantes, as quais levam em consideração os nós com atributos semelhantes. A existência e descoberta de nós com atributos semelhantes está fortemente relacionada ao conceito de homofilia da rede. A homofilia consiste na tendência de nós com atributos semelhantes se conectarem entre si [McPherson et al. 2001]. Em cenários reais, a homofilia tende a ser bastante presente, visto que há uma preferência entre os usuários por manterem relacionamentos com outros usuários que possuem características semelhantes às suas. A Figura 4.9 apresenta um exemplo de um grafo com atributos nos nós. O grafo em questão é composto por 7 nós, ou usuários, e 3 atributos, os quais são: gênero (*M*: masculino; *F*: feminino), idade e altura, sendo o primeiro atributo do tipo categórico e os demais do tipo numérico. Por fim, as arestas presentes na rede conectam dois usuários que são amigos entre si.

No entanto, devido às características inerentes às redes sociais com atributos, os modelos de privacidade diferencial para redes sociais apresentados anteriormente não são suficientemente robustos para capturar essas propriedades e, simultaneamente, prover as garantias de PD. Por mais que um uma rede social ponderada, apresentada na Seção 4.5.3, possa ser vista como uma rede social com atributos nas arestas, visto que os atributos são de natureza numérica, estes apresentam funções diferentes. Em um grafo ponderado, a informação da aresta, também chamada de peso, tem uma importância diferente de acordo com o seu valor. Geralmente, quanto maior o peso, maior a importância da aresta.

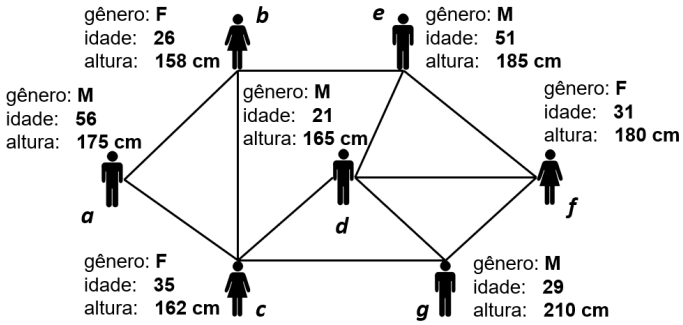


Figura 4.9: Grafo com atributos (gênero, idade e altura) associados aos nós, ou usuários. As arestas representam os relacionamentos entre usuários.

Entretanto, em um grafo com atributos nas arestas, um atributo de natureza numérica desempenhará a função de atributo categórico. Em outras palavras, dadas duas arestas com atributos numéricos de valores diferentes, ambas terão a mesma importância, mesmo que uma aresta tenha um valor muito mais elevado, ou inferior, em relação à outra aresta. Portanto novos modelos de PD para redes sociais com atributos precisam ser propostos a fim de prover as devidas garantias de PD. A seguir, são apresentadas as noções de PD mais utilizadas para redes sociais com atributos nas arestas e nos nós, respectivamente.

Proposta por [Liu et al. 2020], a *attribute-wise* privacidade diferencial (*attribute-wise* PD) consiste na noção de privacidade mais aplicada para grafos com atributos nas arestas. Sua definição é apresentada abaixo:

Definição 14 (ε-*Attribute-wise* Privacidade Diferencial)**.** *Um mecanismo M satisfaz ε-attribute-wise privacidade diferencial se, para qualquer par de grafos com atributos nas arestas vizinhos G e G' diferindo em um atributo e todas as arestas relacionadas ao atributo, e para qualquer saída possível $O \subseteq Range(M)$,*

$$\Pr[M(G) = O] \leq \exp(\varepsilon) \times \Pr[M(G') = O]. \tag{13}$$

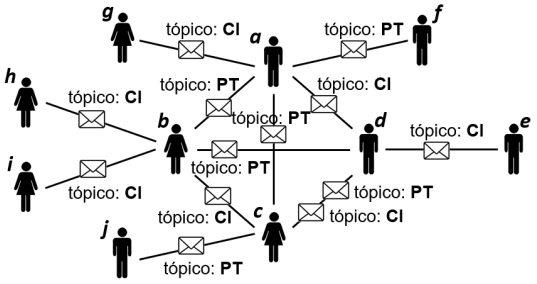
A Figura 4.10 apresenta um exemplo de grafo G com atributos nas arestas e dois possíveis grafos vizinhos G'_1 e G'_2 obtidos a partir de G, conforme a Definição 14. A Figura 4.10a apresenta o grafo original G, enquanto as Figuras 4.10b e 4.10c apresentam os grafos vizinhos G'_1 e G'_2 obtidos, respectivamente, a partir da remoção dos atributos "CI" e "PT" do grafo G.

Por sua vez, [Jorgensen et al. 2016] propôs uma nova noção de privacidade para redes sociais com atributos nos nós, denominada *edge-adjacent attributed* privacidade diferencial (*edge-adjacent attributed* PD). Sua definição é apresentada abaixo:

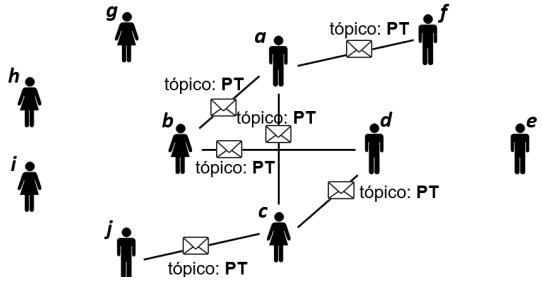
Definição 15 (*Edge-adjacent attributed graphs*)**.** *Dois grafos com atributos nos nós G e G' são ditos edge-adjacent (ou vizinhos) se diferirem em uma única aresta ou nos atributos associados a um único nó.*

Definição 16 (ε-*Edge-adjacent attributed* Privacidade Diferencial)**.** *Um mecanismo M satisfaz ε-edge-adjacent attributed privacidade diferencial se, para qualquer par de grafos edge-adjacent G e G', e para qualquer saída possível $O \subseteq Range(M)$,*

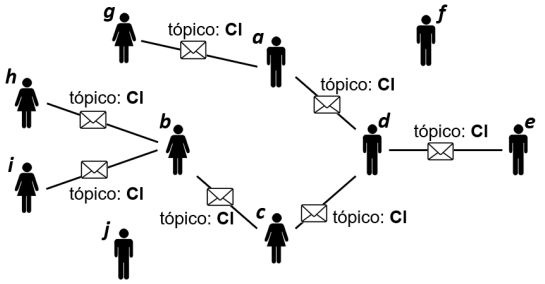
$$\Pr[M(G) = O] \leq \exp(\varepsilon) \times \Pr[M(G') = O]. \tag{14}$$



(a) Rede social G.



(b) Rede social vizinha G'_1 obtida a partir de G após a remoção das arestas com atributo "CI".



(c) Rede social vizinha G'_2 obtida a partir de G após a remoção das arestas com atributo "PT".

Figura 4.10: Exemplo de uma rede social com atributos nas arestas e suas redes sociais vizinhas de acordo a noção *attribute-wise* PD.

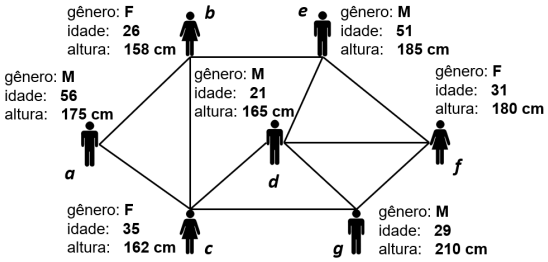
A Figura 4.11 apresenta um exemplo de uma rede social com atributos nos nós G e duas possíveis redes sociais vizinhas G'_1 e G'_2 obtidas a partir de G, conforme a Definição 16. A Figura 4.11a apresenta a rede social G, enquanto as Figuras 4.11b e 4.11c apresentam as redes sociais vizinhas G'_1 e G'_2, respectivamente. Na Figura 4.11b, a rede foi obtida após a remoção da aresta (c, g) da rede G, enquanto que a rede da Figura 4.11c foi obtida após a modificação dos atributos idade e altura do usuário c da rede G.

4.5.5. Definições Complementares de Privacidade Diferencial para Redes Sociais

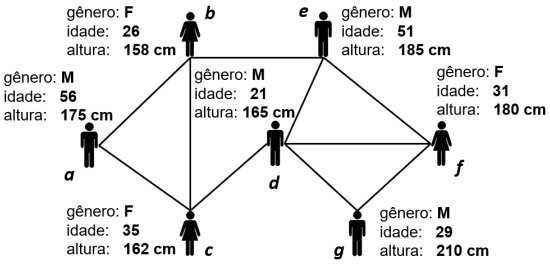
Nas seções anteriores, foram apresentadas várias noções de PD para redes sociais sob diferentes contextos: *edge*-PD, *node*-PD, *edge-weight* PD e *attributed*-PD, como as principais noções no que diz respeito a prover garantias de PD em redes sociais. No entanto, existem diversas noções adicionais de PD que ainda não encontraram uma aplicação muito difundida e são, na sua maioria, derivadas das noções apresentadas anteriormente. Mais detalhes sobre algumas definições e variações adicionais de PD são apresentadas nesta seção.

***Out-link* Privacidade Diferencial:** A primeira dessas novas noções de PD para redes sociais é a *out-link* privacidade diferencial (*out-link* PD) [Task and Clifton 2014]. Para esse contexto, são consideradas redes sociais direcionadas, onde é possível distinguir as arestas que entram e saem dos nós. Sob essa noção, dois grafos são considerados vizinhos se todas as arestas de saída de um nó arbitrário forem adicionadas, ou removidas. A definição formal de *out-link* PD é apresentada abaixo:

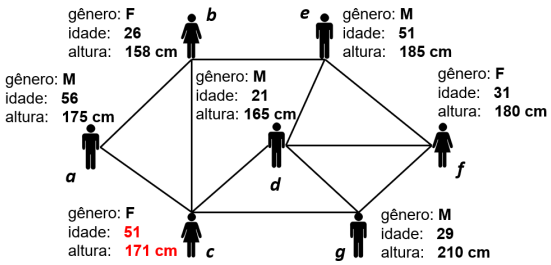
Definição 17 (*Out-link* Privacidade Diferencial)**.** *Dois grafos direcionados são ditos vizinhos se diferirem em todos os out-links (arestas que saem de um nó) de um nó arbitrário.*



(a) Rede social G.



(b) Rede social vizinha G'_1 obtida a partir de G após a remoção da aresta (c, g).



(c) Rede social vizinha G'_2 obtida a partir de G após a modificação dos atributos idade e altura do usuário c.

Figura 4.11: Exemplo de uma rede social com atributos nos nós e suas redes sociais vizinhas de acordo a noção *edge-adjacent attributed* PD.

As garantias de privacidades providas pela *out-link* PD são estritamente mais fracas que as garantias da *node*-PD. Entretanto, em muitos cenários, a *out-link* PD é comparável à *edge*-PD. Sob essa noção de PD, um invasor mal-intencionado não seria capaz de determinar se um usuário contribuiu com seus dados para a construção da rede social com suas respectivas arestas de saída. Em cenários reais, por exemplo, um usuário pode recusar amizades. Por sua vez, outros usuários podem alegar que são amigos de um determinado usuário, mas este último pode negar que essas amizades existem. Dessa forma, os autores argumentam que a *out-link* PD simplifica o cálculo da sensibilidade e reduz a magnitude do ruído adicionado, permitindo consultas que seriam inviáveis sob as definições de *edge*-PD.

***Graph*-Privacidade Diferencial:** A *graph*-privacidade diferencial (*graph*-PD) [Mueller et al. 2022] consiste em uma outra noção de PD para o contexto de redes sociais. Diferente das outras noções apresentadas, a *graph*-PD é uma noção de privacidade mais recente, a qual se aplica tanto à análise de dados em redes sociais, quanto à redes neurais para grafos (GNNs). No entanto, por ser uma noção bem mais recente, a mesma não foi tão explorada ainda. A sua definição formal é apresentada abaixo:

Definição 18 (*Graph*-Privacidade Diferencial)**.** *Dois multigrafos são ditos vizinhos se diferirem em um único grafo, obtido através da adição ou remoção de um gráfico inteiro.*

Apesar de ser uma noção de privacidade recente, a *graph*-PD possui uma grande versatilidade, adaptando-se a diversos contextos diferentes. Dentre os quais a mineração de padrões frequentes e o treinamento de GNNs se destacam em multigrafos. Em resumo, um multigrafo consiste em uma grafo no qual múltiplas arestas entre os mesmos nós são permitidas, de maneira que dois nós podem ser conectados por mais de uma aresta. É

importante destacar que as noções de PD aplicadas à redes sociais não se limitam às que foram apresentadas nesta seção.

4.6. Mecanismos de Privacidade Diferencial para Redes Sociais

Esta seção apresenta uma visão geral dos métodos e abordagens que podem ser aplicados na análise de dados em redes sociais utilizando privacidade diferencial.

4.6.1. Análise Privada para Publicação de Redes Sociais e Redes Sociais Aleatórias

O compartilhamento diferencialmente privado de grafos completos tem sido um segmento de pesquisa estudado extensivamente nos últimos anos. A principal vantagem dessa abordagem é que ela é independente do tipo de análise, de maneira que é possível realizar qualquer tipo de consulta estatística sobre o grafo disponibilizado. Neste cenário, o modelo Pygmalion [Sala et al. 2011] foi proposto com o objetivo de disponibilizar a topologia do grafo sob as garantias de privacidade do modelo *edge*-PD através da extração da estrutura detalhada do grafo em uma versão privada de um dK-grafo [Mahadevan et al. 2006] para, então, gerar um grafo sintético. Posteriormente, melhorias foram propostas na utilidade do dK-grafo por meio de uma melhor calibração do ruído através da sensibilidade *smooth* [Wang and Wu 2013]. Os autores primeiro derivaram do grafo original parâmetros como correlações de grau, e os utilizaram no modelo de dK-grafo, garantindo a privacidade diferencial das arestas nos parâmetros aprendidos para gerar grafos sintéticos. Nos últimos anos, foi desenvolvido um *framework* baseado em microagregação para a anonimização de grafos, denominado *dK-microaggregation*, o qual reduz a magnitude do ruído ao adicionar uma etapa de microagregação ao dK-grafo antes de adicionar o ruído de Laplace [Iftikhar et al. 2020]. Este *framework* é ilustrado na Figura 4.12.

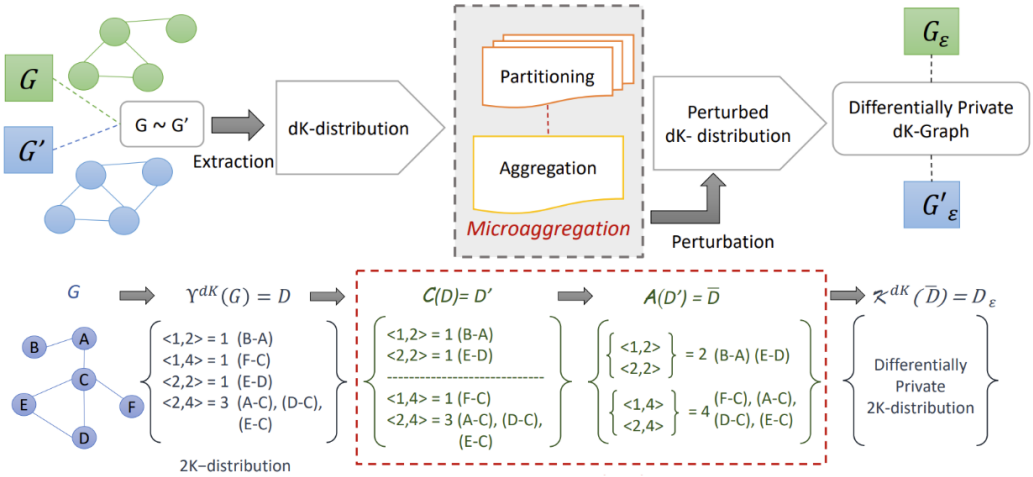


Figura 4.12: *Framework* baseado em microagregação para a anonimização de redes sociais [Iftikhar et al. 2020].

Através de uma abordagem diferente [Xiao et al. 2014], um novo modelo para representação de grafos, denominado *Hierarchical Random Graph* (HRG) [Clauset et al. 2006], foi proposto para a publicação de dados de grafos. Os autores do modelo observaram que, ao estimar as probabilidades de conexão entre nós, a escala de ruído imposta

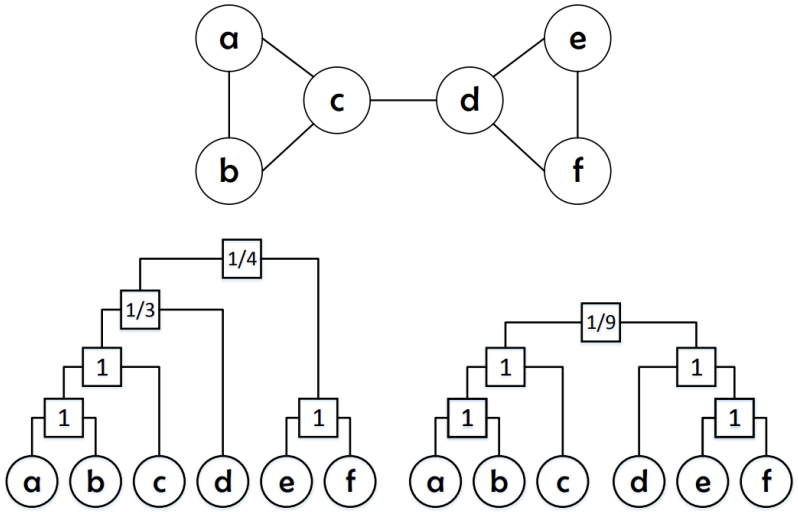


Figura 4.13: Exemplo de um grafo original e dois dendrogramas [Xiao et al. 2014].

pela privacidade diferencial poderia ser reduzida. Modelos clássicos de grafos são construídos considerando as arestas observadas, enquanto o HRG utiliza as probabilidades de conexão entre vértices para formar dendrogramas. Um exemplo de dendrogramas a partir de um grafo de entrada é ilustrado na Figura 4.13.

Adicionalmente, algumas abordagens focam na perturbação da matriz de adjacência original, o que permite representar um grafo e adotar estratégias de perturbação de matriz. O trabalho em [Chen et al. 2014] apresenta um mecanismo de exploração e reconstrução baseado em densidade (DER) para compartilhar a matriz de adjacência de um grafo. No entanto, tanto o HRG quanto o DER possuem complexidade quadrática em relação ao número de nós. Por sua vez, o algoritmo *Top-m Filter* (TmF) [Nguyen et al. 2015] foi proposto com o objetivo de solucionar os problemas de escalabilidade existentes em trabalhos anteriores. O método consiste em adicionar ruído de Laplace a cada célula da matriz de adjacência e utiliza uma ideia semelhante ao algoritmo *High-pass Filter* [Cormode et al. 2012] para evitar a materialização da matriz de adjacência ruidosa.

Os trabalhos mencionados acima utilizam *edge*-PD como modelo de privacidade diferencial para disponibilizar grafos inteiros e aleatórios. No entanto, devido às dificuldades em obter mecanismos privados que proveem alta utilidade de dados após a publicação do grafo completo, há evidências de apenas um único trabalho recente [Jian et al. 2021] relacionado ao modelo de privacidade *node*-PD quando comparado ao modelo de privacidade *edge*-PD. Os autores propõem um algoritmo de perturbação de nós para alcançar o modelo *node*-PD adicionando e removendo nós aleatoriamente. Primeiro, remove-se aleatoriamente cada nó no grafo de entrada de forma independente com probabilidade p. Em seguida, gera-se um número aleatório k, que segue uma distribuição geométrica com probabilidade de sucesso q. Logo, adiciona-se aleatoriamente k nós ao grafo, de modo que cada um desses k nós seja conectado a todos os nós existentes com uma probabilidade de 0,5. Após essa etapa, o número de nós nos grafos resultantes será indistinguível se os grafos de entrada forem grafos vizinhos.

4.6.2. Privacidade em Contagem de Subgrafos

Contagens de subgrafos são outras estatísticas amplamente estudadas no âmbito da análise de dados em grafos. Consultas dessa natureza realizam contagens sobre o número de vezes que uma determinada estrutura de subgrafo aparece em um grafo. Dentre os subgrafos mais comuns, podemos citar os triângulos, as estrelas e os cliques. Nesse contexto, [Karwa et al. 2011] estendeu os resultados da sensibilidade *smooth* [Nissim et al. 2007] para publicar consultas de k-estrelas e k-triângulos de maneira diferencialmente privada. Especificamente, os autores apresentaram um algoritmo para calcular a sensibilidade *smooth* do número de k-estrelas em um grafo de entrada e utilizaram uma abordagem diferente, baseada na sensibilidade local, a fim de fornecer um algoritmo diferencialmente privado para o compartilhamento de contagens de subgrafos. Outra técnica utilizada, a função *ladder* [Zhang et al. 2015] é utilizada para obter uma alta precisão de contagem de subgrafos com complexidades de tempo eficientes. Essa abordagem combina efetivamente o conceito de sensibilidade local à distância t, do *framework* de sensibilidade *smooth* com o mecanismo exponencial, de maneira a permitir, também, a contagem de k-cliques em um grafo.

Nos últimos anos, uma nova noção de privacidade diferencial, denominada *Decentralized Differential Privacy* (DDP) [Sun et al. 2019], foi apresentada juntamente com uma técnica para publicar, de maneira privada, algumas estatísticas de grafos, tais como triângulos, caminhos *three-hop* e contagens de k-cliques através da privacidade diferencial local, ao passo em que garante as propriedades impostas pela DDP. Recentemente, os autores em [Imola et al. 2021] apresentaram novos algoritmos para contagem de subgrafos com privacidade diferencial local que utilizam uma rodada adicional de interação entre os usuários e o curador de dados. Os autores melhoraram os resultados recentes tanto para consultas de contagem de triângulos quanto de k-estrelas.

4.6.3. Privacidade em Histogramas de Graus

Outra estatística de grafo bastante estudada é a sequência de graus de um grafo (*degree sequence*). Em resumo, a sequência de graus consiste em uma lista que contém os graus, ou seja, número de conexões, de cada um dos nós do grafo, geralmente ordenada de maneira decrescente. A sequência de graus torna-se uma estatística bastante robusta quando aplicada, principalmente, em análises envolvendo as distribuições dos graus de um grafo.

Sob o modelo de *edge*-PD, [Hay et al. 2009] propõem uma técnica baseada em inferência de restrições para disponibilizar sequências de graus via mecanismos de privacidade diferencial. Os autores adaptaram a definição de PD para dados estruturados em grafos e foram os primeiros a introduzir a noção de *edge*-PD. O método proposto baseia-se em adicionar ruído às contagens de grau de cada nó. Especificamente, para cada nó, uma pequena quantidade de ruído de Laplace é adicionada à sua contagem de grau verdadeira. Os autores também realizaram uma etapa de pós-processamento nas respostas ruidosas para inferir um resultado mais preciso. Um problema com essa inferência é que a sequência de graus pode não ser realizável, ou seja, não ser possível construir um grafo sem laços ou múltiplas arestas entre o mesmo par de vértices com a sequência de graus retornada pelo mecanismo diferencialmente privado. Para superar isso, [Karwa and Slavković 2012] introduziram uma etapa de otimização após a inferência de restrições. Os

autores propuseram uma abordagem inovadora para gerar grafos sintéticos que satisfazem a privacidade diferencial utilizando um modelo probabilístico baseado na sequência de graus realizável, ou seja, um vetor que descreve a distribuição de graus de um grafo. Eles também incluíram uma etapa adicional de pós-processamento para garantir que a sequência de graus liberada seja realizável.

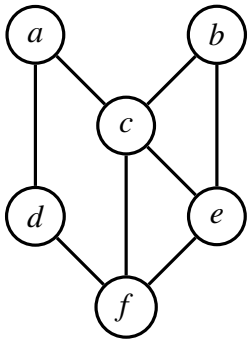
Já no cenário de privacidade diferencial de nós (*node*-PD), os autores em [Kasiviswanathan et al. 2013] discutem vários algoritmos diferencialmente privados voltados para a análise da histogramas de graus. A ideia principal por trás de suas técnicas é projetar o grafo de entrada no conjunto de grafos com um grau máximo abaixo de um determinado limite, denominado θ. No entanto, o desafio dessa abordagem é que a operação de projeção pode ser muito sensível a uma mudança de um único nó no grafo original.

Por outro lado, os autores em [Day et al. 2016] propõem duas abordagens baseadas, respectivamente, em agregação e histograma cumulativo para publicar a distribuição de graus. Ambas as abordagens adotam um novo método de projeção de grafos que é baseado em um processo de adição de arestas. Os autores provaram que publicar um histograma de graus a partir do grafo projetado tem sensibilidade $2\theta + 1$, e publicar um histograma cumulativo de graus tem sensibilidade $\theta + 1$. Esse processo transforma um grafo em um grafo limitado a θ graus. A escolha ideal de θ depende tanto do conjunto de dados quanto do orçamento de privacidade. Em ambas as abordagens, os autores também utilizam uma etapa adicional de pós-processamento para melhorar a precisão dos histogramas. A Figura 4.14 mostra um exemplo de projeção de grafos, para $\theta = 3$ (Figura 4.14b) e para $\theta = 2$ (4.14c), respectivamente.

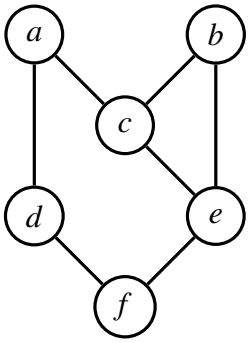
4.6.4. Análise Privada de Medidas de Centralidade

Na análise de dados em redes sociais, estatísticas mais complexas surgem como uma necessidade fundamental, possibilitando a mineração de topologias complexas e, também, a compreensão das interações e relacionamentos entre os indivíduos presentes na rede. Nesse contexto de estatísticas mais complexas, aparecem as medidas de centralidade, as quais são bastante relevantes em mensurar a importância dos nós em um grafo. Tais informações podem ser amplamente aplicadas em análises que envolvem partes de uma rede que necessitam de maiores atenções ou cuidados.

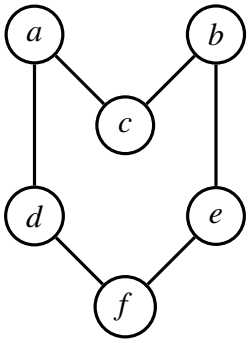
Nesse contexto, o método *PrivateEBC* [Roohi et al. 2019] foi proposto visando resolver o problema relacionado à centralidade de intermediação (*betweenness centrality*) dos nós quando informações relevantes sobre as arestas são disseminadas. Os autores desenvolvem um protocolo que permite calcular a centralidade de intermediação egocêntrica de um nó em um grafo de maneira privada, onde as informações relevantes sobre as arestas estão distribuídas entre duas partes que não se confiam mutuamente, como dois provedores de telecomunicações. Uma outra estratégia que adapta a noção de sensibilidade local para o contexto não numérico [Farias et al. 2020,Farias et al. 2023] é proposta a fim de desenvolver um mecanismo genérico para dados categóricos e disponibilizar informações de centralidade de intermediação utilizando privacidade diferencial. O trabalho busca estender os princípios da privacidade diferencial, tradicionalmente aplicados a dados numéricos, para contextos onde os dados são de natureza não numérica, oferecendo uma abordagem que assegura a privacidade enquanto lida com esse tipo de informação.



(a) Rede social G.



(b) Exemplo de rede social projetada G'_1 com $\theta = 3$.



(c) Exemplo de rede social projetada G'_2 com $\theta = 2$.

Figura 4.14: Exemplo de uma rede social e suas possíveis projeções para diferentes valores de θ.

Por outro lado, o trabalho em [Laeuchli et al. 2022] apresenta limites inferiores e superiores com base na abordagem de sensibilidade *smooth* para centralidade de autovetor, de Laplace e de proximidade. Os autores investigam como a adição de ruído para preservar a privacidade afeta a precisão das medidas de centralidade. No entanto, os autores concluem que a abordagem de sensibilidade *smooth* é inviável ou impraticável.

4.6.5. Detecção de Comunidades de Maneira Privada

Diferentemente das medidas de centralidade, onde o foco de atenção maior era sobre a importância dos indivíduos, uma técnica de detecção de comunidades busca identificar similaridades sobre grupos de indivíduos. Em resumo, a detecção de comunidades consiste em técnicas aplicadas com o objetivo de identificar grupos em redes, geralmente complexas, de acordo com as propriedades estruturais da rede em questão. A expectativa é de que, dado que um grafo pode ser dividido em diversos conjuntos de nós disjuntos, os nós internos de cada um desses conjuntos são densamente conectados entre si.

Nesse contexto, o método LouvainDP [Nguyen et al. 2016], um algoritmo popular para detecção de comunidades, é utilizado como base para a adaptação sob as restrições da privacidade diferencial. Um exemplo do algoritmo de Louvain é demonstrado na Figura 4.15. Para que o algoritmo seja diferencialmente privado, os autores adicionam ruído de Laplace aos pesos das arestas gerados pelo algoritmo de Louvain. A vantagem dessa

estratégia é que ela é linear, em termos de complexidade de algoritmos, em relação ao número de arestas do grafo original. Além disso, os autores incluem uma análise detalhada das dificuldades inerentes à aplicação da privacidade diferencial aos algoritmos de detecção de comunidades.

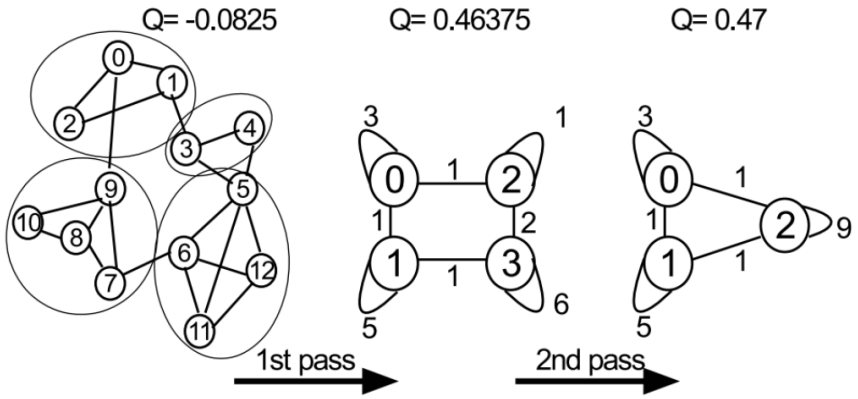


Figura 4.15: Exemplo de detecção de comunidades com o método de Louvain.

Já os autores em [Mohamed et al. 2022] desenvolvem métodos que permitem a detecção de comunidades em redes utilizando o modelo de bloco estocástico (*Stochastic Block Model - SBM*), enquanto garantem a privacidade diferencial. O estudo explora como diferentes configurações do modelo de bloco estocástico afetam o desempenho dos métodos de detecção de comunidades sob privacidade diferencial. Além disso, o mecanismo XOR [Ji et al. 2021] apresenta resultados promissores na detecção de comunidades dada a topologia da rede. Esse novo mecanismo é projetado para fornecer respostas privadas a consultas sobre dados binários e dados em formato de matriz (grafos). O mecanismo perturba os bits dos dados originais, aplicando a operação XOR com um vetor de ruído. Durante uma consulta, os bits perturbados são utilizados para calcular a resposta. A operação XOR garante que a privacidade dos indivíduos é preservada, mesmo que os dados perturbados sejam analisados. Os autores estendem o mecanismo para dados em formato de matriz, onde cada elemento da matriz é perturbado de maneira semelhante aos bits binários, aplicando a operação XOR com um vetor de ruído específico para matrizes.

Por sua vez, o algoritmo DPCD (*Differentially Private Community Detection*) [Ji et al. 2019] consiste em um algoritmo capaz de proteger tanto a topologia da rede quanto os atributos dos nós na detecção de comunidades em redes sociais. A estratégia baseia-se em um modelo probabilístico generativo que realiza a detecção de comunidades resolvendo um problema de máxima verossimilhança, com garantias de privacidade diferencial. Em particular, os autores dividem o problema de máxima verossimilhança em subproblemas convexos, cada um lidando com os relacionamentos e os atributos de um usuário específico. Para proteger os relacionamentos privados de cada usuário, a função objetivo referente às suas relações é perturbada por um ruído injetado com uma distribuição de probabilidade projetada. Para proteger a privacidade dos atributos dos usuários, cada usuário é obrigado a gerar ruído independentemente, enquanto a soma desses ruídos satisfaz a distribuição projetada.

4.6.6. Análise Privada de Caminhos Mínimos e Distâncias

Quando os grafos possuem arestas ponderadas, os modelos de privacidade diferencial para arestas e nós mencionados anteriormente podem não oferecer garantias de privacidade adequadas. Em vez disso, um ajuste mais apropriado é adaptar a definição de privacidade diferencial no contexto de grafos ponderados.

Sealfon [Sealfon 2016] propôs fundamentos teóricos para considerar a privacidade diferencial dos pesos em um grafo. Seu estudo visa compartilhar caminhos mínimos ponderados entre pares de nós e compartilhar distâncias aproximadas entre todos os pares de nós sem revelar informações sensíveis sobre os pesos das arestas. Para o problema de liberar caminhos mais curtos de forma privada, o autor apresentou um limite inferior robusto baseado em reconstrução, mostrando que não é possível disponibilizar um caminho mínimo entre um par de vértices com erro adicionado melhor que $\Omega(|V|)$, sob privacidade diferencial. Esse limite inferior é obtido ao reduzir o problema de reconstruir muitas das linhas de um banco de dados ao problema de encontrar um caminho com baixo erro. O autor também mostrou que um algoritmo que utiliza o mecanismo de Laplace chega perto de atingir esse limite. Já considerando o problema de compartilhar caminhos mais curtos ponderados entre todos os pares de nós com privacidade diferencial, o autor argumentou que técnicas padrão produzem erro de $O(|V|log|V|)/\varepsilon$ para cada consulta diferencialmente privada. Por outro lado, ele obteve algoritmos melhorados para duas classes especiais de grafos: (1) árvores e (2) grafos com pesos de arestas limitados.

[Fan and Li 2022] revisitou o problema de disponibilizar distâncias aproximadas entre todos os pares de nós de maneira privada e melhorou os resultados de Sealfon. Os autores primeiro dividiram uma árvore em caminhos pesados disjuntos. Em um caminho pesado, cada nó não-folha seleciona um ramo, ou seja, a aresta para o filho que tem a maior profundidade. As arestas selecionadas formam um caminho pesado. A Figura 4.16 exemplifica três caminhos pesados na árvore de entrada.

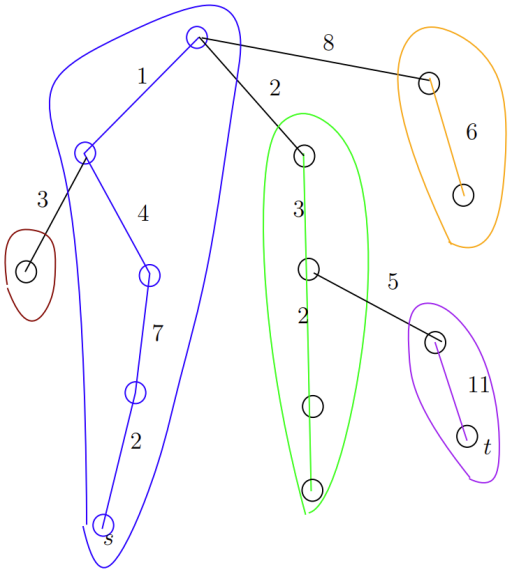


Figura 4.16: Um exemplo de uma árvore particionada em três caminhos pesados. Uma cor única é atribuída a cada caminho pesado [Fan and Li 2022].

Como a árvore é decomposta em um conjunto de caminhos, algumas arestas podem não estar incluídas em nenhum dos caminhos pesados produzidos durante este método. Já o trabalho em [Brito et al. 2023] estabelece uma nova definição de grafos vizinhos considerando tanto a topologia do grafo quanto os pesos das arestas como informações privadas. Os autores fornecem uma solução escalável de perturbação de grafos ponderados e introduzem tanto uma abordagem global quanto uma abordagem utilizando privacidade diferencial local. O modelo de privacidade utilizado é o *edge-weight* privacidade diferencial (*edge-weight* PD). Os autores obtém resultados significativos em várias métricas relacionadas a caminhos mínimos e distâncias, como soma dos pesos totais das arestas, média de caminhos mínimos, soma dos pesos das arestas adjacentes a um nó, entre outros. A Figura 4.17 ilustra um exemplo de como essa abordagem funciona.

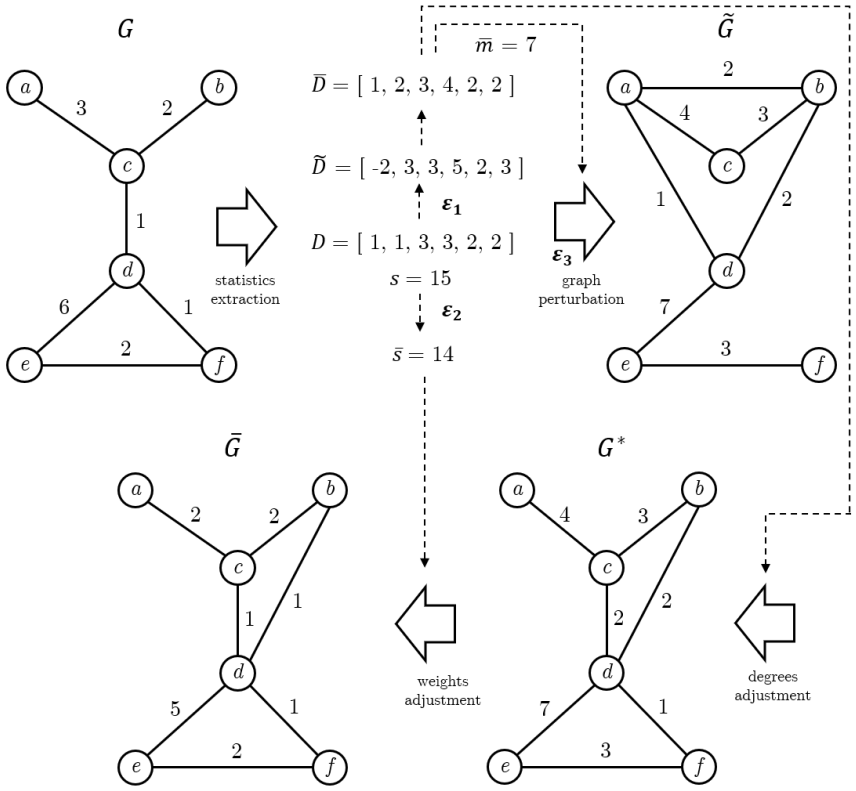


Figura 4.17: Exemplo de compartilhamento de um grafo ponderado utilizando *edge-weight* PD.

4.6.7. Privacidade em Redes Neurais para Redes Sociais

O recente sucesso das redes neurais impulsionaram pesquisas sobre reconhecimento de padrões e mineração de dados. Diversas tarefas de aprendizado de máquina, como detecção de objetos, tradução automática e reconhecimento de fala, ganharam bastante atenção através dos paradigmas de aprendizado profundo, como as redes neurais convolucionais (RNC), redes neurais recorrentes (RNR) e codificadores automáticos. No entanto, os métodos de aprendizado profundo se mostram adequados para identificar padrões sobre dados euclidianos, como imagens, textos e vídeos, inviabilizando sua aplicação sobre dados gerados a partir de domínios não euclidianos, como as estruturas de grafos, as quais

são utilizadas para modelar complexos relacionamentos e interdependências entre entidades. Por conta das limitação dos modelos existentes, recentemente, as redes neurais para grafos (GNN) receberam atenção significativa devido, principalmente, à sua capacidade de capturar relacionamentos complexos e dependências entre os nós de um grafo. As GNN pertencem à uma classe de redes neurais que podem ser aplicadas sobre estruturas de grafos. Dentre os tipos de aplicações mais conhecidas, pode-se mencionar tarefas relacionadas à classificação de nós e grafos, visualização de grafos, predição de conexões, agrupamento de grafos, dentre outras.

No entanto, por se tratarem de um conjunto de aplicações que fazem uso direto dos dados, questões quanto à privacidade dos indivíduos tornam-se novamente aparentes. Nesse contexto, algumas técnicas diferencialmente privadas começaram a surgir. Dentre as primeiras técnicas a surgirem, o *framework* PrivGnn permite a proteção de dados sensíveis mesmo após a publicação de um modelo de redes neurais para grafos. Por sua vez, [Daigavane et al. 2021] propôs um esquema de amostragem de vizinhança de grafos ao passo em que assegura o modelo de privacidade *node*-PD. No entanto, todos os trabalhos recém mencionados reforçam a privacidade somente durante a etapa de treinamento e/ou liberação do modelo. Esse fato pode colocar informações acerca dos indivíduos em sérios riscos, caso a parte interessada seja maliciosa. Diante dessa limitação, recentemente, foi proposto o *framework* RGNN [Bhaila et al. 2023], o qual é baseado na reconstrução e aprendizado de redes neurais para grafos e garante a privacidade dos nós do grafo através de privacidade diferencial local.

4.7. Conclusão

A análise de dados privados em redes sociais é um desafio essencial no contexto atual, onde a preservação da privacidade dos usuários é crucial. Realizar análises de dados de maneira privada é fundamental para manter a privacidade dos usuários e proteger informações sensíveis contra acessos indevidos. A evolução dos modelos de privacidade reflete o progresso nesta área, começando com modelos sintáticos e avançando para abordagens mais robustas como a privacidade diferencial.

Os modelos sintáticos, como a anonimização e a generalização, inicialmente tentaram proteger a privacidade substituindo, ou ocultando, informações que identificavam unicamente os usuários. No entanto, a reidentificação através de ataques complexos demonstrou as limitações desses modelos, revelando a necessidade de técnicas mais avançadas. Nesse contexto, a privacidade diferencial emergiu como como o novo padrão para a proteção de dados, oferecendo uma abordagem matemática robusta para quantificar e limitar os riscos de privacidade.

A privacidade diferencial apresenta duas configurações principais: a global e a local. Na configuração global, o ruído é adicionado aos dados, ou resultados das análises, através de um curador centralizado, o qual detém a posse dos dados. Em contrapartida, na configuração local, o controle dos dados é colocado sob responsabilidade dos próprios usuários, garantindo que os dados sejam protegidos no momento da coleta, antes mesmo de serem enviados ao curador de dados, o qual é considerado não confiável. Cada uma das configurações apresenta suas vantagens e desvantagens, sendo a escolha dependente do contexto e das necessidades específicas de privacidade e precisão das análises.

Além das considerações de privacidade, é importante reconhecer a diversidade das redes sociais. Existem desde as redes sociais mais simples, constituídas apenas por nós e arestas, onde cada nó representa um usuário e cada aresta representa um relacionamento entre os usuários. No entanto, também existem redes sociais bem mais complexas como, por exemplo, as redes sociais ponderadas, onde as conexões têm pesos, os quais representam a intensidade, ou frequência, das interações. Um outro tipo de rede são as redes sociais com atributos, as quais as arestas, ou nós, possuem características adicionais que podem influenciar nas análises. Assim, a variedade e complexidade das redes sociais exigem abordagens flexíveis e adaptativas para a análise de dados privada. Portanto, avançar na implementação de modelos de privacidade eficazes, como a privacidade diferencial, em diferentes tipos de redes sociais é crucial para garantir que a análise de dados possa ser realizada de maneira segura, garantindo a privacidade dos indivíduos enquanto se extraem *insights* valiosos.

No entanto, apesar da vasta literatura existente sobre privacidade diferencial para redes sociais, ainda destaca-se a carência de ferramentas de código aberto acessíveis para a implementação dessas técnicas. Essa lacuna não só limita a aplicação prática das metodologias discutidas, como também representa uma barreira para pesquisadores e desenvolvedores que buscam integrar privacidade diferencial em suas análises em redes sociais de maneira eficiente e replicável. Além disso, diversas estatísticas valiosas, como as medidas de centralidade, ainda carecem de versões diferencialmente privadas. Essa ausência aponta para uma rica área de pesquisa em aberto, onde a adaptação e desenvolvimento de métodos que garantam a privacidade diferencial para essas estatísticas podem trazer significativos avanços, tanto teóricos, quanto práticos.

Outro aspecto essencial é a necessidade de alinhar as expectativas e interpretações dos usuários em relação à privacidade diferencial aplicada em redes sociais. O orçamento de privacidade ε é muitas vezes considerado contraintuitivo, além de que a compreensão das diferentes abordagens de privacidade, seja a nível de aresta, nó ou atributo, são áreas que exigem maior clareza na sua explicação. Além disso, o desempenho computacional apresenta-se como um fator limitante na implementação de técnicas de privacidade diferencial para análises em redes sociais. As técnicas atuais, na grande maioria das vezes, exigem recursos computacionais consideráveis, o que pode limitar sua aplicabilidade em cenários com grandes volumes de dados, ou em tempo real. Por fim, estender a privacidade diferencial para ambientes de redes sociais dinâmicas, com fluxo de dados contínuos (*streaming*), permanece sendo uma área aberta para pesquisa.

Em resumo, a realização de análise privada de dados em redes sociais é um campo dinâmico que demanda inovação contínua para equilibrar a utilidade dos dados com a necessidade de proteger a privacidade dos usuários. No entanto, apesar dos avanços significativos realizados nos modelos de privacidade, diversos desafios e oportunidades de pesquisa permanecem em aberto. O desenvolvimento de novas técnicas, juntamente com a adaptação contínua a diferentes estruturas de redes sociais, é essencial para um futuro onde a privacidade e a análise de dados possam coexistir harmoniosamente.

Agradecimentos

Este trabalho foi parcialmente financiado pela Lenovo, como parte do seu investimento em Pesquisa e Desenvolvimento (P&D) de acordo com a Lei de Informática. Os autores também agradecem ao financiamento fornecido pela CAPES, sob os processos de número 88881.189723/2018-01 e 88882.454571/2019-01, e pelo CNPq, sob o processo de número 316729/2021-3.

Referências

[Abdulhamid et al. 2014] Abdulhamid, S. M., Ahmad, S., Waziri, V. O., and Jibril, F. N. (2014). Privacy and national security issues in social networks: the challenges. *arXiv preprint arXiv:1402.3301*.

[Acharya et al. 2019] Acharya, J., Sun, Z., and Zhang, H. (2019). Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1120–1129. PMLR.

[Allardyce 2012] Allardyce, C. S. (2012). *Fat chemistry: The science behind obesity*. Royal Society of Chemistry.

[Alsmadi and Alhami 2015] Alsmadi, I. and Alhami, I. (2015). Clustering and classification of email contents. *Journal of King Saud University-Computer and Information Sciences*, 27(1):46–57.

[Baden et al. 2009] Baden, R., Bender, A., Spring, N., Bhattacharjee, B., and Starin, D. (2009). Persona: an online social network with user-defined privacy. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pages 135–146.

[Bassily and Smith 2015] Bassily, R. and Smith, A. (2015). Local, private, efficient protocols for succinct histograms. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 127–135.

[Bhaila et al. 2023] Bhaila, K., Huang, W., Wu, Y., and Wu, X. (2023). Local differential privacy in graph neural networks: a reconstruction approach. *arXiv preprint arXiv:2309.08569*.

[Bloch et al. 2023] Bloch, F., Jackson, M. O., and Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, 61(2):413–453.

[Boyd and Ellison 2007] Boyd, D. M. and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230.

[Brazil 2018] Brazil (2018). Lei Geral de Proteção de Dados Pessoais. `http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm`. Acesso em: 22 de maio de 2024.

[Brito 2023] Brito, F. T. (2023). *Differentially private release of count-weighted graphs*. PhD thesis, Universidade Federal do Ceará.

[Brito et al. 2023] Brito, F. T., Farias, V. A., Flynn, C., Majumdar, S., Machado, J. C., and Srivastava, D. (2023). Global and local differentially private release of count-weighted graphs. *Proceedings of the ACM on Management of Data*, 1(2):1–25.

[Brito and Machado 2017] Brito, F. T. and Machado, J. C. (2017). Preservação de privacidade de dados: Fundamentos, técnicas e aplicações. *Jornadas de atualização em informática*, pages 91–130.

[Brito et al. 2024] Brito, F. T., Mendonça, A. L. C., and Machado, J. C. (2024). A differentially private guide for graph analytics. In *Proceedings 27th International Conference on Extending Database Technology, EDBT 2024, Paestum, Italy, March 25 - March 28*, pages 850–853. OpenProceedings.org.

[Brito et al. 2015] Brito, F. T., Neto, A. C. A., Costa, C. F., Mendonça, A. L., and Machado, J. C. (2015). A distributed approach for privacy preservation in the publication of trajectory data. In *Proceedings of the 2nd Workshop on Privacy in Geographic Information Collection and Analysis*, pages 1–8.

[Chaabane et al. 2012] Chaabane, A., Acs, G., Kaafar, M. A., et al. (2012). You are what you like! information leakage through users' interests. In *Proceedings of the 19th annual network & distributed system security symposium (NDSS)*. Citeseer.

[Chen et al. 2022] Chen, L., Han, K., Xiu, Q., and Gao, D. (2022). Graph clustering under weight-differential privacy. In *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, pages 1457–1464. IEEE.

[Chen et al. 2014] Chen, R., Fung, B., Yu, P. S., and Desai, B. C. (2014). Correlated network data publication via differential privacy. *The VLDB Journal*, 23(4):653–676.

[Clauset et al. 2006] Clauset, A., Moore, C., and Newman, M. E. (2006). Structural inference of hierarchies in networks. In *ICML Workshop on Statistical Network Analysis*, pages 1–13. Springer.

[Cook and Holder 2006] Cook, D. J. and Holder, L. B. (2006). *Mining graph data*. John Wiley & Sons.

[Cormode et al. 2018] Cormode, G., Jha, S., Kulkarni, T., Li, N., Srivastava, D., and Wang, T. (2018). Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, pages 1655–1658.

[Cormode et al. 2012] Cormode, G., Procopiuc, C., Srivastava, D., and Tran, T. T. (2012). Differentially private summaries for sparse data. In *Proceedings of the 15th International Conference on Database Theory*, pages 299–311.

[da Costa Filho and Machado 2023] da Costa Filho, J. S. and Machado, J. C. (2023). FELIP: A local differentially private approach to frequency estimation on multidimensional datasets. In *Proceedings 26th International Conference on Extending Database*

Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023, pages 671–683. OpenProceedings.org.

[Daigavane et al. 2021] Daigavane, A., Madan, G., Sinha, A., Thakurta, A. G., Aggarwal, G., and Jain, P. (2021). Node-level differentially private graph neural networks. *arXiv:2111.15521*.

[Dalenius 1977] Dalenius, T. (1977). Towards a methodology for statistical disclosure control.

[Day et al. 2016] Day, W.-Y., Li, N., and Lyu, M. (2016). Publishing graph degree distribution with node differential privacy. In *Proceedings of the 2016 International Conference on Management of Data*, pages 123–138.

[Dey et al. 2012] Dey, R., Tang, C., Ross, K., and Saxena, N. (2012). Estimating age privacy leakage in online social networks. In *2012 proceedings ieee infocom*, pages 2836–2840. IEEE.

[Duchi et al. 2013] Duchi, J. C., Jordan, M. I., and Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE.

[Dwork 2006] Dwork, C. (2006). Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pages 1–12. Springer.

[Dwork et al. 2006] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer.

[Dwork et al. 2014] Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

[Erlingsson et al. 2014] Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. ACM.

[European Commission 2018] European Commission (2018). 2018 reform of EU data protection rules. `https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf`. Acesso em: 22 de maio de 2024.

[Fan and Li 2022] Fan, C. and Li, P. (2022). Distances release with differential privacy in tree and grid graph. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2190–2195. IEEE.

[Farias et al. 2020] Farias, V. A., Brito, F. T., Flynn, C., Machado, J. C., Majumdar, S., and Srivastava, D. (2020). Local dampening: differential privacy for non-numeric queries via local sensitivity. *Proceedings of the VLDB Endowment*, 14(4):521–533.

[Farias et al. 2023] Farias, V. A., Brito, F. T., Flynn, C., Machado, J. C., Majumdar, S., and Srivastava, D. (2023). Local dampening: Differential privacy for non-numeric queries via local sensitivity. *The VLDB Journal*, pages 1–24.

[Garfinkel et al. 2018] Garfinkel, S. L., Abowd, J. M., and Powazek, S. (2018). Issues encountered deploying differential privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 133–137. ACM.

[Ghosh et al. 2009] Ghosh, A., Roughgarden, T., and Sundararajan, M. (2009). Universally utility-maximizing privacy mechanisms. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 351–360.

[Gong et al. 2014] Gong, N. Z., Talwalkar, A., Mackey, L., Huang, L., Shin, E. C. R., Stefanov, E., Shi, E., and Song, D. (2014). Joint link prediction and attribute inference using a social-attribute network. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(2):1–20.

[Hay et al. 2009] Hay, M., Li, C., Miklau, G., and Jensen, D. (2009). Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining*, pages 169–178. IEEE.

[Hsu et al. 2014] Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B. C., and Roth, A. (2014). Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE.

[Iftikhar et al. 2020] Iftikhar, M., Wang, Q., and Lin, Y. (2020). dk-microaggregation: Anonymizing graphs with differential privacy guarantees. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 191–203. Springer.

[Imola et al. 2021] Imola, J., Murakami, T., and Chaudhuri, K. (2021). Locally differentially private analysis of graph statistics. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 983–1000.

[Ji et al. 2021] Ji, T., Li, P., Yilmaz, E., Ayday, E., Ye, Y., and Sun, J. (2021). Differentially private binary-and matrix-valued data query: an xor mechanism. *Proceedings of the VLDB Endowment*, 14(5):849–862.

[Ji et al. 2019] Ji, T., Luo, C., Guo, Y., Ji, J., Liao, W., and Li, P. (2019). Differentially private community detection in attributed social networks. In *Asian Conference on Machine Learning*, pages 16–31. PMLR.

[Jian et al. 2021] Jian, X., Wang, Y., and Chen, L. (2021). Publishing graphs under node differential privacy. *IEEE Transactions on Knowledge and Data Engineering*.

[Jiang et al. 2021] Jiang, H., Pei, J., Yu, D., Yu, J., Gong, B., and Cheng, X. (2021). Applications of differential privacy in social network analysis: A survey. *IEEE transactions on knowledge and data engineering*, 35(1):108–127.

[Jorgensen et al. 2016] Jorgensen, Z., Yu, T., and Cormode, G. (2016). Publishing attributed social graphs with formal privacy guarantees. In *Proceedings of the 2016 international conference on management of data*, pages 107–122.

[Karwa et al. 2011] Karwa, V., Raskhodnikova, S., Smith, A., and Yaroslavtsev, G. (2011). Private analysis of graph structure. *PVLDB*, 4(11):1146–1157.

[Karwa and Slavković 2012] Karwa, V. and Slavković, A. B. (2012). Differentially private graphical degree sequences and synthetic graphs. In *International Conference on Privacy in Statistical Databases*, pages 273–285. Springer.

[Kasiviswanathan et al. 2013] Kasiviswanathan, S. P., Nissim, K., Raskhodnikova, S., and Smith, A. (2013). Analyzing graphs with node differential privacy. In *Theory of Cryptography Conference*, pages 457–476. Springer.

[Kaytoue et al. 2017] Kaytoue, M., Plantevit, M., Zimmermann, A., Bendimerad, A., and Robardet, C. (2017). Exceptional contextual subgraph mining. *Machine Learning*, 106:1171–1211.

[Kenthapadi et al. 2019] Kenthapadi, K., Mironov, I., and Thakurta, A. (2019). Privacy-preserving data mining in industry. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1308–1310. ACM.

[Kossinets and Watts 2006] Kossinets, G. and Watts, D. J. (2006). Empirical analysis of an evolving social network. *science*, 311(5757):88–90.

[Laeuchli et al. 2022] Laeuchli, J., Ramírez-Cruz, Y., and Trujillo-Rasua, R. (2022). Analysis of centrality measures under differential privacy models. *Applied Mathematics and Computation*, 412:126546.

[Li et al. 2023] Li, L., Zhao, Y., Luo, S., Wang, G., and Wang, Z. (2023). Efficient community search in edge-attributed graphs. *IEEE Transactions on Knowledge and Data Engineering*.

[Li et al. 2006] Li, N., Li, T., and Venkatasubramanian, S. (2006). t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*, pages 106–115. IEEE.

[Li et al. 2016] Li, N., Lyu, M., Su, D., and Yang, W. (2016). Differential privacy: From theory to practice. *Synthesis Lectures on Information Security, Privacy, & Trust*, 8(4):1–138.

[Li et al. 2017] Li, X., Yang, J., Sun, Z., and Zhang, J. (2017). Differential privacy for edge weights in social networks. *Security and Communication Networks*, 2017.

[Liu et al. 2020] Liu, Z., Huang, L., Xu, H., Yang, W., and Wang, S. (2020). Privag: Analyzing attributed graph data with local differential privacy. In *2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 422–429. IEEE.

[Machanavajjhala et al. 2007] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3–es.

[Mahadevan et al. 2006] Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Computer Communication Review*, 36(4):135–146.

[McPherson et al. 2001] McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

[McSherry and Talwar 2007] McSherry, F. and Talwar, K. (2007). Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pages 94–103. IEEE.

[Mendonça et al. 2023] Mendonça, A. L., Brito, F. T., and Machado, J. C. (2023). Privacy-preserving techniques for social network analysis. In *Anais Estendidos do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 174–178. SBC.

[Mislove et al. 2010] Mislove, A., Viswanath, B., Gummadi, K. P., and Druschel, P. (2010). You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260.

[Mitchell 2017] Mitchell, J. S. (2017). Shortest paths and networks. In *Handbook of discrete and computational geometry*, pages 811–848. Chapman and Hall/CRC.

[Mohamed et al. 2022] Mohamed, M. S., Nguyen, D., Vullikanti, A., and Tandon, R. (2022). Differentially private community detection for stochastic block models. In *International Conference on Machine Learning*, pages 15858–15894. PMLR.

[Mueller et al. 2022] Mueller, T. T., Paetzold, J. C., Prabhakar, C., Usynin, D., Rueckert, D., and Kaissis, G. (2022). Differentially private graph neural networks for whole-graph classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Narayanan and Shmatikov 2008] Narayanan, A. and Shmatikov, V. (2008). Robust deanonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE.

[Narayanan and Shmatikov 2009] Narayanan, A. and Shmatikov, V. (2009). Deanonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 173–187. IEEE.

[Near and Abuah 2021] Near, J. P. and Abuah, C. (2021). Programming differential privacy. *URL: https://uvm*.

[Nergiz et al. 2007] Nergiz, M. E., Atzori, M., and Clifton, C. (2007). Hiding the presence of individuals from shared databases. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pages 665–676.

[Nguyen et al. 2015] Nguyen, H. H., Imine, A., and Rusinowitch, M. (2015). Differentially private publication of social graphs at linear cost. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 596–599. IEEE.

[Nguyen et al. 2016] Nguyen, H. H., Imine, A., and Rusinowitch, M. (2016). Detecting communities under differential privacy. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*, pages 83–93.

[Nissim et al. 2007] Nissim, K., Raskhodnikova, S., and Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84.

[Pinot et al. 2018] Pinot, R., Morvan, A., Yger, F., Gouy-Pailler, C., and Atif, J. (2018). Graph-based clustering under differential privacy. *arXiv preprint arXiv:1803.03831*.

[Prell 2011] Prell, C. (2011). *Social network analysis: History, theory and methodology*. Sage.

[Qian et al. 2016] Qian, J., Li, X.-Y., Zhang, C., and Chen, L. (2016). De-anonymizing social networks and inferring private attributes using knowledge graphs. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE.

[Ribeiro et al. 2021] Ribeiro, P., Paredes, P., Silva, M. E., Aparicio, D., and Silva, F. (2021). A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

[Roohi et al. 2019] Roohi, L., Rubinstein, B. I., and Teague, V. (2019). Differentially-private two-party egocentric betweenness centrality. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 2233–2241. IEEE.

[Sala et al. 2011] Sala, A., Zhao, X., Wilson, C., Zheng, H., and Zhao, B. Y. (2011). Sharing graphs using differentially private graph models. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 81–98.

[Samoylenko et al. 2023] Samoylenko, I., Aleja, D., Primo, E., Alfaro-Bittner, K., Vasilyeva, E., Kovalenko, K., Musatov, D., Raigorodskii, A. M., Criado, R., Romance, M., et al. (2023). Why are there six degrees of separation in a social network? *Physical Review X*, 13(2):021032.

[Sealfon 2016] Sealfon, A. (2016). Shortest paths and distances with differential privacy. In *Proceedings of the 35th Symposium on Principles of Database Systems*, pages 29–41.

[Shah et al. 2016] Shah, N., Beutel, A., Hooi, B., Akoglu, L., Gunnemann, S., Makhija, D., Kumar, M., and Faloutsos, C. (2016). Edgecentric: Anomaly detection in edge-attributed networks. In *2016 IEEE 16Th international conference on data mining workshops (ICDMW)*, pages 327–334. IEEE.

[Shi et al. 2016] Shi, C., Li, Y., Zhang, J., Sun, Y., and Philip, S. Y. (2016). A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 29(1):17–37.

[Silva et al. 2017] Silva, R. R. C., Leal, B. C., Brito, F. T., Vidal, V. M., and Machado, J. C. (2017). A differentially private approach for querying rdf data of social networks. In *Proceedings of the 21st International Database Engineering & Applications Symposium*, pages 74–81.

[Sun et al. 2019] Sun, H., Xiao, X., Khalil, I., Yang, Y., Qin, Z., Wang, H., and Yu, T. (2019). Analyzing subgraph statistics from extended local views with decentralized differential privacy. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 703–717.

[Sweeney 2002] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570.

[Task and Clifton 2014] Task, C. and Clifton, C. (2014). What should we protect? defining differential privacy for social network analysis. In *State of the Art Applications of Social Network Analysis*, pages 139–161. Springer.

[Wang et al. 2013] Wang, C.-D., Lai, J.-H., and Philip, S. Y. (2013). Neiwalk: Community discovery in dynamic content-based networks. *IEEE transactions on knowledge and data engineering*, 26(7):1734–1748.

[Wang and Long 2019] Wang, D. and Long, S. (2019). Boosting the accuracy of differentially private in weighted social networks. *Multimedia Tools and Applications*, 78(24):34801–34817.

[Wang et al. 2017] Wang, T., Blocki, J., Li, N., and Jha, S. (2017). Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 729–745.

[Wang and Wu 2013] Wang, Y. and Wu, X. (2013). Preserving differential privacy in degree-correlation based graph generation. *Transactions on data privacy*, 6(2):127.

[Xiao et al. 2014] Xiao, Q., Chen, R., and Tan, K.-L. (2014). Differentially private network data release via structural inference. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 911–920.

[Ye and Barg 2018] Ye, M. and Barg, A. (2018). Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676.

[Zhang et al. 2015] Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., and Xiao, X. (2015). Private release of graph statistics using ladder functions. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 731–745.