Games and Economic Behavior 105 (2017) 255–275

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Privacy-constrained network formation ☆

Daron Acemoglu [a,*], Ali Makhdoumi [b], Azarakhsh Malekian [c], Asuman Ozdaglar [b]

[a] Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States
[b] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, United States
[c] Rotman School of Management, University of Toronto, Toronto, ON, M5S 3E6, Canada

ARTICLE INFO

Article history:
Received 11 August 2015
Available online 9 August 2017

JEL classification:
D85

Keywords:
Network formation
Networks
Privacy

ABSTRACT

We study the effects of privacy concerns on social network formation. Each individual decides which others to form links with. Links bring direct benefits from friendship but also lead to the sharing of information via a percolation process. Privacy concerns are modeled as a disutility that the individual suffers as a result of her private information being acquired by others. We specify conditions under which pure-strategy equilibria exist and characterize both pure-strategy and mixed-strategy equilibria. The resulting equilibrium networks feature clustered connections and homophily. Clustering emerges because if player a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, making it less costly for a to befriend c. Homophily emerges because small additional benefits of friendship within a group make linkages and thus information sharing within that group more likely, further increasing the likelihood within-group links.

1. Introduction

With the increasing volume of information-sharing in online platforms, privacy has become a central concern. Many individuals are willing to take costly actions in order to prevent platforms, retailers, advertisers as well as acquaintances from having access to their private information (Varian, 2009). Though many online platforms, including Facebook, have taken steps to alleviate these concerns, privacy is likely to become even a more important issue in the years to come.

There is relatively little work, however, on how privacy concerns impact online behavior. In this paper, we take a first step by considering a network formation game in the presence of information leakage over the network, which is costly for individuals. The network in question can stand for the friendship or connections network over a social media platform or as an abstract representation of online trading activity. Thus the insights from our analysis should apply both to social media and to online commercial activities.

In our model, each individual decides to form directed links to others. Links have heterogeneous benefits (e.g., an individual receives benefits from befriending others in a social media platform). Once formed, these links also transmit information, however. For example, a friendship link over a social media platform inevitably involves some information sharing, while

☆ We thank anonymous reviewers and various numerous seminar and conference participants for useful suggestions. We gratefully acknowledge financial support from the Toulouse Network with Information Technology and Army Research Office, ARO MURI W911NF-12-1-0509.

* Corresponding author.
E-mail addresses: daron@mit.edu (D. Acemoglu), makhdoum@mit.edu (A. Makhdoumi), azarakhsh.malekian@rotman.utoronto.ca (A. Malekian), asuman@mit.edu (A. Ozdaglar).

http://dx.doi.org/10.1016/j.geb.2017.08.001
0899-8256/

online trades will necessarily give information to the user's trading partner about his or her preferences. More important than the direct transmission of this information is indirect transmission over the network: the relevant information can travel not only to one's friend but also to friends of friends, etc. This makes the cost of loss of privacy for an individual a function of the equilibrium network (which other friendships have formed in equilibrium).

Though the social interactions captured by our model are potentially complex, the setup is relatively parsimonious. It consists of a matrix of benefits of direct links, a cost of loss of privacy (the cost of an individual's information being observed by each of other agents in the network), and percolation process for the travel of information over the network.

We first characterize properties of the equilibrium network. Our first result identifies an endemic problem of non-existence of pure-strategy equilibria. This can be best understood by considering relationship between three agents. If player b has formed a link to player c, this will discourage player a from forming a link with player b, because any information shared with b now risks traveling to c. This in turn, encourages player c to form a link with player a. Finally, as c has formed a link with a, b would disconnect from c, resulting in a contradiction.

We also establish that a sufficient and necessary condition for the existence of pure-strategy equilibria is that the matrix of benefits of direct links is such that high benefit subset of edges have no directed cycle. In particular, we show that if this condition holds, then a pure-strategy Nash equilibrium always exists and if this condition does not hold, there exists a matrix of benefits consistent with high benefit subset of edges for which no pure-strategy Nash equilibrium exists.

As a final characterization result, we also establish an interesting phase transition in pure-strategy equilibria as we vary the ability of information transmission. As this transmission probability increases, there are two opposing forces: first, with higher transmission probability, an agent's information is likely to reach any other agent she is indirectly connected to, and this discourages connections. But secondly, and in contradiction to the first force, this greater transmission probability also implies that the cost of connecting directly to such an agent is lower, thus encouraging greater connections. We show that the resolution of these two opposing forces implies that until the transmission probability reaches a critical threshold, the equilibrium network is sparse, but as this critical threshold is reached, the equilibrium becomes a collection of densely-connected components with clustering coefficient one.

Our other sets of results concern the patterns of connections that occur (in pure or mixed-strategy equilibria). Our first major result shows that equilibrium networks feature clustered connections. This pattern emerges because if a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, thus making it less costly for a to befriend c. Second, we also show that the equilibrium network features homophily. The reason for this is that even an infinitesimal advantage in terms of direct benefits of friendship within a group makes linkages within that group more likely, in turn making information travel within that group and reducing the cost of making further within-group links due to loss of privacy. This increases the likelihood of further within-group links.

Though there is relatively little work on how privacy affects individual decisions in online platforms and social media settings, there are several other large and growing literature to which our paper relates.[1] First, our work is part of a large literature on network economics and endogenous social network formation. This includes works on network formation such as Jackson and Wolinsky (1996); Bala and Goyal (1997, 2000); Galeotti et al. (2006); Tardos and Wexler (2007); Galeotti and Goyal (2010); Dutta and Mutuswami (1997); Jackson and Watts (2002); Celis and Mousavifar (2015); Jackson (2003). It also includes works on network games such as Bramoullé and Kranton (2007); Galeotti et al. (2010); Ballester et al. (2006); Jackson and Rogers (2007); Currarini et al. (2009); Acemoglu et al. (2015), as well as works on contagion in networks such as Morris (2000); Blume et al. (2013); Elliott et al. (2014); Acemoglu et al. (2016). In particular, Celis and Mousavifar (2015) studies a model in which agents form directed links to other agents with certain cost and gain utility from indirect connections to other agents and Currarini et al. (2009) develops a model of friendship formation where individuals receive type-dependent benefits from friendship, and explains the emergence of homophily in friendships and how this varies with group size. Closely related are Kleinberg et al. (2001); Fabrikant et al. (2003); Blume et al. (2013), and Kleinberg and Ligett (2013) which study a model of the trade-off between the benefits received from sharing information and the cost of indirect sharing of information, though both their models and results are very different from ours. In particular, our paper relates to that of Kleinberg and Ligett (2013) in that both include a model of the trade-off between the benefits of sharing information with friends, and the risks that additional gossiping will propagate it to people with whom one is not on friendly terms. However, the stochastic process governing gossip, the equilibrium notion (Nash equilibrium versus stable equilibrium), and the utility of agents are different. In addition, the two papers focus on fundamentally different questions. Kleinberg and Ligett (2013) study the computability of stable networks and its inefficiency compared to social welfare maximizing network. In contrast, we study the properties of equilibrium network such as clustering and homophily.

Second, a large sociology and network literature emphasizes the regularity of triadic closure. The basic principle is that if two people in a social network have a friend in common, then there is an increased likelihood that they become friends themselves at some point in the future (see Rapoport, 1953 and Easley and Kleinberg, 2010). This pattern can be detected either by verifying triadic closure properties or focusing on various network statistics that provide information on this, such as the *clustering coefficient* (see e.g., Newman, 2003, 2004; Watts and Strogatz, 1998, Watts and Strogatz, 1998, and

[1] See Dwork and Roth (2013), Liang et al. (2009), Kearns et al. (2016), and Lori (2012) for surveys on various aspects of privacy concerns, Calzolari and Pavan (2006), Laudon (1996), Taylor (2004), Hui and Png (2006), Posner (1981), and Acquisti et al. (2016) for examples of economics papers discussing privacy-related issues, and Samuelson (2000), Westin (1968), Stigler (1980), Hirshleifer (1980), Magi (2011), and Newell (2011) for certain legal aspects of privacy.

Fagiolo, 2007). Evidence on these patterns is provided in, among others, Medus and Dorso (2013), Kossinets and Watts (2006), Albert and Barabási (2002), Davidsen et al. (2002), Holme and Kim (2002), and Vázquez (2003).

Third, there is also a similarly large literature on the second key pattern generated by privacy-constrained network formation: homophily. Homophily, defined as the tendency of people to associate with others similar to themselves, is observed in many social networks, ranging from friendships to marriages to business relationships, and is based on a variety of characteristics and attributes, including ethnicity (see e.g. Fong and Isajiw, 2000 and Baerveldt et al., 2004), race, age, gender, religion, education level, profession, political affiliation, and other attributes (see e.g. Lazarsfeld et al., 1954, Blau, 1997, Blalock, 1982, Marsden, 1988, Marsden, 1987, and the survey by McPherson et al., 2001). Various different explanations for homophily have been proposed in, among others, Moody (2001), Patacchini and Zenou (2016), Currarini et al. (2009), and Fowler et al. (2009).

2. Model

We consider a set $\mathcal{V} = \{1, \ldots, n\}$ of agents interested in forming friendship links with each other.[2] In choosing their links, agents tradeoff the benefit from direct links with the cost of privacy loss due to leakage of information through indirect links. Each agent makes a decision about connecting to other agents, i.e., agent i chooses a strategy $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$, where $x_{ij} \in \{0, 1\}$ represents whether agent i is connected to agent j or not (we use the convention that $x_{ii} = 0$ for all $i \in \mathcal{V}$). We assume that the decision $x_{ij} = 1$ results in a directed friendship link from i to j, implying i shares her information with j and receives friendship benefits from it but not necessarily the other way around, i.e., x_{ji} need not be equal to x_{ij}. This means that if agent i shares her information with j (e.g., to get advise on a matter), agent j does not have to share her information with i. Though most prior literature (see Jackson, 2005) considers undirected models of friendship in the context of network formation, in several settings links and friendships are not always on equal footing and can be more fruitfully modeled as directed links (see e.g. Bala and Goyal, 1997 and Celis and Mousavifar, 2015 for models with directed links). For instance, a friendship might involve one party, individual a, sharing information with another, b, either in the course of social interactions or to receive some advice, while b does not share any information in return. In the context of social media, the amount of information shared between friends and connected individuals is again often asymmetric. Finally, another application of these ideas would be to other online interactions, such as individuals using a website, platform or service, and uploading information in the process (e.g., likes and dislikes or credit card information).[3]

Given node i's decision x_{ij} with respect to agent j, agent i derives a benefit $v_{ij}x_{ij}$ from her friendship with agent j, where $v_{ij} \geq 0$ is a parameter that captures the value i has for her friendship with j. We collect all v_{ij}'s in an $n \times n$ matrix $\mathbf{V} = [v_{ij}]_{i,j \in \mathcal{V}}$ and refer to it as the *valuation matrix* (we use the convention that $v_{ii} = 0$ for all $i \in \mathcal{V}$). We also collect the strategy of all agents in an $n \times n$ matrix $\mathbf{x}$ where $[\mathbf{x}]_{ij} = x_{ij}$. We also use $\mathbf{x}_{-i} = (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$ to denote the strategy profile of all agents other than i. A given strategy profile $\mathbf{x}$ induces a (directed) graph among agents $G_\mathbf{x} = (\mathcal{V}, E_\mathbf{x})$, where $E_\mathbf{x}$ is the set of directed edges given by

$$E_\mathbf{x} = \left\{(i, j) \in \mathcal{V} \times \mathcal{V} \mid x_{ij} = 1\right\},$$

where $(i, j) \in E_\mathbf{x}$ implies that in the graph $G_\mathbf{x}$ there is a direct link from i to j. The matrix $\mathbf{x}$ can be viewed as the adjacency matrix of the graph $G_\mathbf{x}$. Next, we define the process for the indirect leakage of information and the utility function of the agents.

Information from agent (node) i leaks to other agents over $G_\mathbf{x}$ according to a gossip process, which is an independent cascade process of the following form (see Kempe et al., 2003 for more details). Information from i reaches each of her out-neighbors $l \in N^{\text{out}}(i)$ with probability 1 at time 1 where $N^{\text{out}}(i) = \{l \in \mathcal{V} : (i, l) \in E_\mathbf{x}\}$. We call individuals *informed about* i (or simply informed) if i's information has reached to them, and *uninformed about* i (or simply uninformed) otherwise. We assume that once a node becomes informed, it will remain so forever. The dynamics of information propagation is as follows: when a node j first becomes informed, say at time t, it becomes a *source* of information. It then stochastically, simultaneously, and independently informs each of its previously uninformed neighbors. In particular, each uninformed neighbor of j, such as $k \in N^{\text{out}}(j)$, becomes informed at time t with probability β, independent of each other and the history of the process (and in this case, k itself becomes a source at time $t + 1$). If an uninformed node is a neighbor to multiple informed sources (i.e., there exists at least two j and j' such that $k \in N^{\text{out}}(j)$ and $k \in N^{\text{out}}(j')$), then it can become informed by either in an order-independent fashion (e.g., if k is uninformed and a neighbor to two sources, then the probability that it will become informed is $1 - (1 - \beta)^2$). Node j that was a source at time t becomes a *non-source* in all subsequent periods. That is, if a neighbor of node j does not become informed at time t, then it will never again become informed via node j.

We refer to $\beta \in [0, 1]$ as the *transmission parameter*. This process essentially assumes that information from i leaks with transmission probability 1 on the neighboring edges and transmission probability β on all other edges independently. It captures the assumption that direct neighbors/friends of i have access to all information of i, while indirect friends may

[2] We use the terms agents, players, and individuals interchangeably.

[3] This last application would require other, relatively straightforward, changes (e.g., considering a bipartite graph of users and online platforms, with information flows in only one direction, and making decisions about which platforms are linked and share information among themselves).

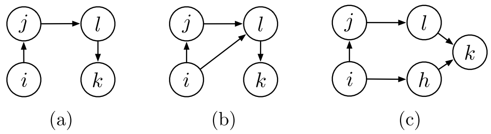


Fig. 1. An example to illustrate the gossip process and the expected aggregate leakage of information.

obtain information of i through a gossip process. Given a graph $G_{\mathbf{x}} = (\mathcal{V}, E_{\mathbf{x}})$ formed by the decision of individuals, we will next present an equivalent description of the gossip process, which is convenient for analysis.

Definition 1 *(Gossip process).* For a given node $i \in \mathcal{V}$, we denote the probability that information from i reaches node j by $\mathbb{P}[i \rightsquigarrow j]$ which is defined as follows. We draw a realized graph $G = (\mathcal{V}, E)$, on which we activate outgoing edges of i on $G_{\mathbf{x}}$ with probability 1 and all other edges of $G_{\mathbf{x}}$ with probability β independently (therefore, we have $E \subseteq E_{\mathbf{x}}$). Information from i will reach node j if and only if there is a directed path of active edges from i to j on this realized graph, in which case we say node j is reachable from node i and write $i \rightsquigarrow j$.[4] The probability that the graph G is realized is

$$P_{i,\mathbf{x}}(G) = \left(\prod_{(i,l)\in E_{\mathbf{x}}} \mathbf{1}\{(i,l) \in E\} \right) \left(\prod_{\substack{(k,k')\in E_{\mathbf{x}},\ k\neq i \\ (k,k')\in E}} \beta \right) \left(\prod_{\substack{(k,k')\in E_X,\ k\neq i \\ (k,k')\notin E}} (1-\beta) \right),$$

where the subscript i and $\mathbf{x}$ denote the dependence of this probability on agent i and $\mathbf{x}$ and $\mathbf{1}\{\cdot\}$ is the indicator function. Therefore, the probability that information from i reaches node j is given by

$$\mathbb{P}[i \rightsquigarrow j] = \sum_{G\in\mathcal{G}} P_{i,\mathbf{x}}(G)\mathbf{1}\{i \rightsquigarrow j\},$$

where $\mathcal{G}$ denotes the set of all possible graphs with the set of nodes $\mathcal{V}$ and the set of edges which is a subset of E_X, i.e.,

$$\mathcal{G} = \{G = (\mathcal{V}, E) \mid E \subseteq E_{\mathbf{x}}\}.$$

As an illustration, consider the network given in Fig. 1a. The only realized graph $G = (V, E)$ for which i's information reaches k is the one with $E = \{(i,j),(j,l),(l,k)\}$, which has probability β^2; thus $\mathbb{P}[i \rightsquigarrow k] = \beta^2$. On the other hand, for the network given in Fig. 1b, there are two realized graphs $G = (V, E)$ for which i's information reaches k (note that (i, j) and (i,l) belong to any realized graph as j and l are neighbors of i). The edge set of these two realized graphs are: (i) $E = \{(i,j),(i,l),(l,k)\}$, which has probability $\beta(1-\beta)$; and (ii) $E = \{(i,j),(i,l),(l,k),(j,l)\}$, which has probability β^2, implying $\mathbb{P}[i \rightsquigarrow k] = \beta(1-\beta) + \beta^2 = \beta$. Finally, for the network given in Fig. 1c, there are five realized graphs $G = (V, E)$ for which i's information reaches k: (i) $E = \{(i,j),(i,h),(h,k)\}$, which has probability $\beta(1-\beta)^2$; (ii) $E = \{(i,j),(i,h),(h,k),(j,l)\}$, which has probability $\beta^2(1-\beta)$; (iii) $E = \{(i,j),(i,h),(h,k),(l,k)\}$; which has probability $\beta^2(1-\beta)$, (iv) $E = \{(i,j),(i,h),(h,k),(j,l),(l,k)\}$, which has probability β^3; and (v) $E = \{(i,j),(i,h),(j,l),(l,k)\}$, which has probability $\beta^2(1-\beta)$. Combining these five events, we have $\mathbb{P}[i \rightsquigarrow k] = \beta(1-\beta)^2 + 3\beta^2(1-\beta) + \beta^3 = \beta + \beta^2 - \beta^3$.

The expected aggregate leakage of information of i, denoted by $\Gamma(i, \mathbf{x})$ is the summation of probabilities $\mathbb{P}[i \rightsquigarrow j]$ over all $j \in \mathcal{V}$. With the convention that for any $i \in \mathcal{V}$, $\mathbb{P}[i \rightsquigarrow i] = 0$, it can be written as

$$\Gamma(i,\mathbf{x}) = \sum_{j\in\mathcal{V}} \mathbb{P}[i \rightsquigarrow j].$$

Note that $\Gamma(i, \mathbf{x})$ is a function of the adjacency matrix $\mathbf{x}$ as well as the transmission parameter β.

The utility function of agent i, denoted by u_i, is given by

$$u_i(\mathbf{x}) = \sum_{j\in\mathcal{V}} x_{ij} v_{ij} - \gamma\, \Gamma(i,\mathbf{x}), \tag{1}$$

where the *cost parameter* $\gamma \geq 0$ captures the tradeoff between the benefit of friendship and the loss of privacy. For a given valuation matrix $\mathbf{V}$, which is assumed to be known by all agents, we define the Nash equilibrium of the complete information game as follows.

[4] Throughout the paper, we use the notation $i \rightsquigarrow j$ to denote that j is reachable from i, and the notation $i \to j$ to show that i has a direct link to j, i.e., $x_{ij} = 1$. We will also use the notation $i \nrightarrow j$ to denote $x_{ij} = 0$.

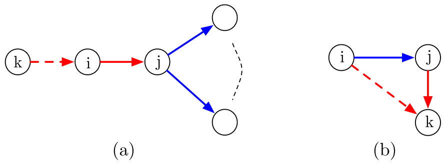


Fig. 2. (a) strategic substitutability. (b) strategic complementarity.

Definition 2 *(Pure-strategy Nash equilibrium).* The set of strategies $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a pure-strategy Nash equilibrium if for all $i \in \mathcal{V}$, we have

$$\mathbf{x}_i \in \underset{\mathbf{y_i} \in \{0,1\}^n}{\operatorname{argmax}} \, u_i(\mathbf{y_i}, \mathbf{x}_{-i}).$$

We refer to the network formed induced by the strategies $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as the *equilibrium network*, where $\mathbf{x}$ shows its adjacency matrix.

We focus on Nash equilibria, rather than pairwise stability as in Jackson and Wolinsky (1996); Blume et al. (2013), and Kleinberg and Ligett (2013) as we wish to focus on each individuals' incentive to form links unilaterally in the context of a directed friendship network (see also Jackson, 2005 for an overview of other solution concepts).[5]

Because, as we show below, pure-strategy Nash equilibria may not always exist, we also consider mixed-strategy Nash equilibria, defined in the usual fashion.

Definition 3 *(Mixed-strategy Nash equilibrium).* The mixed strategy $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, where σ_i is a probability measure over $\{0, 1\}^n$, is a mixed-strategy Nash equilibrium if for any i, we have

$$u_i(\sigma_i, \boldsymbol{\sigma}_{-i}) \geq u_i(\mathbf{y}_i, \boldsymbol{\sigma}_{-i}), \text{ for any } \mathbf{y}_i \in \{0, 1\}^n,$$

where $\boldsymbol{\sigma}_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$.

Finally, when we turn to the analysis of homophily, there will sometimes be additional, unintuitive equilibria. One way of eliminating these is to consider strong (pure-strategy) Nash equilibria, which also test for deviations by coalitions (we will also establish that our other results are valid regardless of whether we use pure-strategy Nash equilibrium or strong Nash equilibrium).

Definition 4 *(Strong Nash equilibrium).* A set of decisions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a strong pure-strategy Nash equilibrium if and only if there exists no *coalition* $S \subseteq \{1, \ldots, n\}$ that has a profitable deviation, i.e., none of the agents of S receives a lower utility after deviation and at least one of the agents of S receives a higher utility.

The analysis of the Nash equilibrium of this game is made complicated by the fact that it features both strategic substitutability and strategic complementarity. In particular, a link from an agent i to agent j generally discourages another agent k from forming a link to i because of increased likelihood of leakage of k's information, introducing an element of strategic substitutes (see Fig. 2a). On the other hand, when there is already a link from i to j, then if j forms a link to k, i would become more likely to form a link to k because her information is already leaked indirectly to k, which is an element of strategic complements (see Fig. 2b).

3. Existence of equilibria

In this section, we study the existence of both pure and mixed-strategy Nash equilibria. When either transmission parameter β or cost parameter γ is zero, a pure-strategy Nash equilibrium always exists.[6] In the rest of this section, we will focus on strictly positive values of transmission parameter $\beta > 0$ and cost parameter $\gamma > 0$.

We first give an example in which a pure-strategy Nash equilibrium does not exist.

[5] This is particularly natural in the case of connections in the context of social media, which are unilateral, directed links. In the case of friendship, we interpret links to represent how much trust an individual puts in an acquaintance or member of broader community, which is again better represented as a directed link decided unilaterally. In other settings, however, joint decisions might be important because a friendship may require participation by both parties.

[6] If transmission parameter $\beta = 0$, for any agent i and j where $v_{ij} \geq \gamma$ we let $x_{ij} = 1$. These decisions clearly form an equilibrium. Similarly, if $\gamma = 0$, then a complete graph is a pure-strategy Nash equilibrium. The cases $\beta = 0$ or $\gamma = 0$ correspond to situations where agents do not face any loss of utility due to privacy breach and there will be no trade-off between the benefit of friendship and the cost of indirect leakage of information.

Example 1. Let $\gamma > 0$ and $\beta > 0$. Consider three agents, denoted by a, b, and c, and let $v_{ab} = v_{bc} = v_{ca} = \gamma(1 + \frac{\beta}{4})$ and $v_{ba} = v_{cb} = v_{ac} = \gamma(1 - \frac{\beta}{2})$. First note that in any pure-strategy Nash equilibrium, we have $x_{ba} = x_{cb} = x_{ac} = 0$. To see this suppose the contrary that $x_{ba} = 1$. Therefore, the utility of b is upper bounded by

$$u_b(\mathbf{x}) \leq \max\{v_{ba} - \gamma, v_{ba} + v_{bc} - 2\gamma\} < 0,$$

which is negative and a profitable deviation for b would be to disconnect from both a and c in order to obtain zero utility. Therefore, an equilibrium network (if exists) belongs to the set of eight possible networks defined by $(x_{ab}, x_{bc}, x_{ca}) \in \{0, 1\}^3$. Next, we will argue that none of them can be an equilibrium network.

- an empty network is not an equilibrium network as a has a profitable deviation which is $x_{ab} = 1$. This deviation would increase her utility from 0 to $v_{ab} - \gamma = \gamma\frac{\beta}{4}$.
- a network with one edge is not an equilibrium network. Without loss of generality (because of symmetry), suppose $x_{ab} = 1$. Player b has a profitable deviation which is $x_{bc} = 1$. This deviation would increase her utility from 0 to $v_{bc} - \gamma = \gamma\frac{\beta}{4}$.
- a network with two edges is not an equilibrium network. Without loss of generality, suppose $x_{ab} = x_{bc} = 1$. Player a has a profitable deviation, i.e., $x_{ab} = 0$. This deviation would increase her utility from $v_{ab} - \gamma(1 + \beta) = -\frac{3}{4}\gamma\beta$ to 0.
- a network with three edges is not an equilibrium network for the same reason as in the previous case.

As suggested in Example 1, if there exists a directed cycle on the set of the edges with valuations higher than the cost γ, then a pure-strategy Nash equilibrium might not exist. We will next establish a necessary and sufficient condition for the existence of a Nash equilibrium in terms of the graph formed by edges whose valuation parameters are "high", denoted by popular-connections which is defined next.

Definition 5 *(Popular-connections graph).* For a given valuation matrix $\mathbf{V}$, a connection (i, j) is called *popular* if its valuation is at least γ, i.e., $v_{ij} \geq \gamma$. We define the popular-connections graph of $\mathbf{V}$ as a directed graph with vertex set $\mathcal{V}$ with an edge between two nodes i and j if and only if (i, j) is popular, i.e., popular-connections graph is the graph $(\mathcal{V}, E_\gamma)$ where

$$E_\gamma = \left\{(i, j) \in \mathcal{V} \times \mathcal{V} \mid v_{ij} \geq \gamma\right\}.$$

We say that the valuation matrix $\mathbf{V}$ is compatible with $(\mathcal{V}, E_\gamma)$, if $(\mathcal{V}, E_\gamma)$ is the popular-connections graph of $\mathbf{V}$.

In words, the popular-connections graph includes edges where the direct benefit of connection, v_{ij}, exceeds the direct cost from loss of privacy, γ. It is also useful to observe that each node would form the connections in this graph if there were no other edges formed or if $\beta = 0$ (because in both cases the total cost from loss of privacy is exactly γ).

In the next theorem, we show that the necessary and sufficient condition for the existence of a pure-strategy Nash equilibrium is indeed the absence of directed cycles of length at least three in popular-connections graph.

Theorem 1. *Let $\beta > 0$ and $\gamma > 0$. Sufficient and necessary conditions for the existence of pure-strategy Nash equilibrium are as follows.*

1. *If the popular-connections graph $(\mathcal{V}, E_\gamma)$ contains a simple directed cycle (of length at least three),[7] then there exists an assignment of valuation matrix $\mathbf{V}$ compatible with $(\mathcal{V}, E_\gamma)$ and a transmission parameter $\bar{\beta}$ such that for all $\beta < \bar{\beta}$, there does not exist a pure-strategy Nash equilibrium.*
2. *If the popular-connections graph $(\mathcal{V}, E_\gamma)$ has no directed cycle, then for any $\mathbf{V}$ compatible with $(\mathcal{V}, E_\gamma)$ there exists a pure-strategy Nash equilibrium. Furthermore, starting from an empty graph ($\mathbf{x}_i = 0$, for all $i \in \mathcal{V}$), the best response dynamics converges to a pure-strategy Nash equilibrium.*[8]

The proof idea of Theorem 1 is as follows. For the first part, similar to Example 1, we show that if the popular-connections graph has a cycle, then there exists a valuation matrix $\mathbf{V}$ for which no pure-strategy Nash equilibrium exists. The proof of the second part of Theorem 1 is constructive. It starts from creating nested sets of nodes R_i, $i \geq 1$, such that for any i the outneighbors of nodes in $R_i \setminus R_{i-1}$ in the graph $(\mathcal{V}, E_\gamma)$ are all in R_{i-1} (we let $R_0 = \emptyset$). This can be done since $(\mathcal{V}, E_\gamma)$ contains no directed cycle.[9] We say all the nodes belonging to set $R_i \setminus R_{i-1}$ have rank i. We then construct a strategy profile by iteratively considering, at each step i, the best response decisions of agents in $\mathcal{V} \setminus R_i$ regarding connecting to nodes in R_i (keeping all previous decisions of other agents fixed). We show that the strategy profile resulting from this construction is an equilibrium. This construction is further illustrated in the following example.

[7] A simple cycle of a graph is a cycle with no repetition of nodes.

[8] Best response dynamics starting from an empty graph is defined as a sequence of decisions denoted by $\mathbf{x}(t)$ at time t updated as follows: (i) $\mathbf{x}(0) = 0$, (ii) for any $t \geq 0$, we let $\mathbf{x}_i(t+1) \in \text{armax}_{\mathbf{y} \in \mathbb{R}^n} u_i(\mathbf{y}, \mathbf{x}_{-i}(t))$ (see Fudenberg and Levine, 1998, Chapter 2).

[9] This is very similar to how a "topological sort" is constructed in directed graphs with no directed cycle (see e.g. Leiserson et al., 2001). A topological sort provides a ranking on the nodes such that if $a \rightsquigarrow b$ in the graph, then a has a higher rank than b.

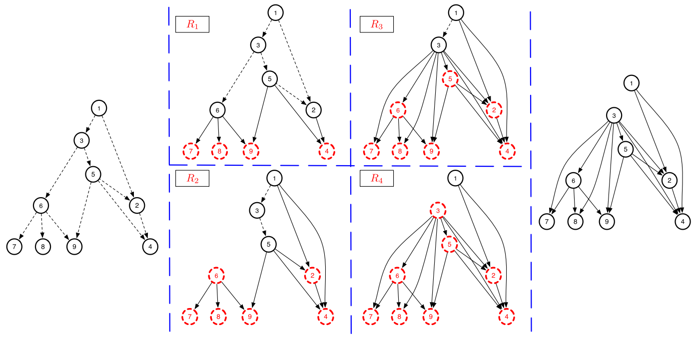


Fig. 3. Construction of pure-strategy Nash equilibrium: each step i shows the set R_i as well as the best responses regarding connections to all nodes in R_i.

Example 2. We consider $(\mathcal{V}, E_\gamma)$ in Fig. 3, where we have 9 nodes. We suppose that all the valuations along the edges of $(\mathcal{V}, E_\gamma)$ are H, all the other valuations are L, and $\beta = 1$, where $L < \gamma < H$, $(H-\gamma)+3(L-\gamma) < 0$, $(H-\gamma)+2(L-\gamma) > 0$, and $2(H-\gamma)+5(L-\gamma) > 0$ (this is guaranteed for instance for $H \in (\frac{9}{4}\gamma, \frac{10}{4}\gamma)$ and $L = \frac{1}{2}\gamma$). We first show the construction of nested sets R_i, $i \geq 1$: we let R_1, the set of rank-1 nodes, be the set of nodes with no outgoing edges in the graph $(\mathcal{V}, E_\gamma)$. In this example we have $R_1 = \{7, 8, 9, 4\}$. For each $i \geq 1$, we update R_i to obtain R_{i+1} by including all the nodes in $\mathcal{V} \setminus R_i$ whose outneighbors in the graph $(\mathcal{V}, E_\gamma)$ are all in R_i. We claim that this procedure stops at some m for which $R_m = \mathcal{V}$.[10] This is because: (i) R_1, i.e., the initial set of nodes with no outgoing edges exists as otherwise the graph $(\mathcal{V}, E_\gamma)$ would have a directed cycle. (ii) at each step i either $R_i = \mathcal{V}$ or we can find R_{i+1} such that $R_{i+1} \setminus R_i \neq \emptyset$ and all out-neighbors of the nodes in $R_{i+1} \setminus R_i$ are in R_i. Because, otherwise each node in $\mathcal{V} \setminus R_i$ would have at least one outgoing edge in $\mathcal{V} \setminus R_i$, showing there exists a directed cycle in the graph $(\mathcal{V}, E_\gamma)$ (among a subset of nodes of $\mathcal{V} \setminus R_i$) which contradicts the assumption. (iii) Finally, the process stops because there are finitely many nodes in the graph. In the example shown in Fig. 3 we have $R_2 = \{7, 8, 9, 4, 2, 6\}$, $R_3 = \{7, 8, 9, 4, 2, 6, 5\}$, $R_4 = \{7, 8, 9, 4, 2, 6, 5, 3\}$, and $R_5 = \mathcal{V}$.

We next define the construction of the strategy profile. At step i we let all nodes in $\mathcal{V} \setminus R_i$ play their best response decisions regarding connecting to nodes in R_i without connecting to other nodes given the current decisions of all other nodes. Note that at step i the decisions of nodes in $\mathcal{V} \setminus R_i$ regarding connecting to nodes in R_i are decoupled. This holds because there is no directed path between any $j, j' \in \mathcal{V} \setminus R_i$ over the edges formed until and including step i (since j and j' only connect to nodes in R_i and there is no edge formed thus far outgoing from R_i). This implies that the connection decisions of j' does not affect j's cost due to loss of privacy and hence its utility. In the example shown in Fig. 3 the best response decision of nodes in $\mathcal{V} \setminus R_1$ regarding nodes in R_1 are $x_{67} = x_{68} = x_{69} = x_{59} = x_{24} = x_{54} = 1$ (this strategy profile is labeled by R_1 in Fig. 3). In the second step, we let all players play their best response decision regarding the nodes in R_2. Here, the decisions are $x_{36} = 0$ (because $(H-\gamma)+3(L-\gamma) < 0$) and $x_{52} = x_{12} = x_{14} = 1$. In the third step, we let all players play their best response decisions regarding nodes in R_3 resulting in the decisions $x_{35} = x_{34} = x_{32} = x_{36} = x_{37} = x_{38} = x_{39} = 1$ because $2(H-\gamma)+5(L-\gamma) > 0$ and $(H-\gamma)+3(L-\gamma) < 0$. In the fourth step, the best response decisions regarding nodes in R_4 is $x_{12} = x_{14} = 1$. The construction stops as $\mathcal{V} \setminus R_5 = \emptyset$. The resulting network depicted in the right hand side of Fig. 3 is an equilibrium network.

To show that this construction leads to an equilibrium network, we will consider, for instance, node 5 and show that it does not have a profitable deviation at the end of the construction. The last decision that node 5 makes is at the second step regarding connections to nodes in $R_2 = \{7, 8, 9, 4, 2, 6\}$ in which she decides to connect to 2, 4, and 9. We next consider the possible deviations for node 5 and show that she would not deviate from her decision at the end of the construction. First, note that node 5 does not make a connection to nodes in $\{1, 3\}$. By construction of the nested sets, there does not exist a direct link on $(\mathcal{V}, E_\gamma)$ from 5 to nodes in $\{1, 3\}$. Moreover, by construction of the strategy profile, there exists no (indirect) path (in the graph obtained at the end of construction) from 5 to nodes in $\{1, 3\}$. These together imply that at the final strategy profile, 5 would receive a negative utility by connecting to nodes in $\{1, 3\}$. Since connection decision of 5 to nodes in R_2 were optimal, it follows that 5 does not have a profitable deviation, showing that the resulting profile is a Nash equilibrium.

Though Theorem 1 shows that pure-strategy Nash equilibria exist when there does not exist a directed cycle in the popular-connections graph, the existence of a directed cycle does not necessarily imply non-existence of pure-strategy Nash equilibria as shown in the next example.

Example 3. Consider a graph with four nodes $\{a, b, c, d\}$ such that $v_{ab} = v_{bc} = v_{cd} = v_{da} = \gamma\left(1 + \frac{\beta}{2}\right)$ and suppose that all other valuations are zero, which implies that the popular-connections graph has a directed cycle of length four. Nevertheless, the following profile is a pure-strategy Nash equilibrium: $x_{ab} = x_{cd} = 1$, and no other connections.

[10] If the $\mathcal{V}$ consist of several subsets with no connection among them, then we apply the same construction procedure on each of those subsets.

Note that if popular-connections graph has no directed cycle, then the pure-strategy Nash equilibrium constructed in Theorem 1 is the unique equilibrium network. We next show that this equilibrium is also a strong Nash equilibrium, and thus it is robust against deviations by coalitions.

Proposition 1. *Let $\beta > 0$ and $\gamma > 0$. If the popular-connections graph has no directed cycle, then the constructed pure-strategy Nash equilibrium in Theorem 1 is also a strong Nash equilibrium.*

This result also follows by considering the aforementioned construction of nested sets and pure-strategy Nash equilibrium. To see this, let us consider a subset of nodes S that has a profitable deviation. We will show that for any $i \geq 1$ no node from R_i can belong to S. First, note that no node from R_1 can belong to set S. This is because any connection that a node from R_1 makes (regardless of other connection decisions) would result in a negative utility. We next show that no node from $R_2 \setminus R_1$ can belong to the set S. This is because any connection from nodes in R_2 to nodes in $\mathcal{V} \setminus R_2$ leads to a negative marginal utility (regardless of other connections). Since nodes in R_2 do not have a path to nodes in $\mathcal{V} \setminus R_2$, the decision of the nodes in $\mathcal{V} \setminus R_2$ is irrelevant for the utility of nodes in R_2. Finally, the decision of nodes in $R_2 \setminus R_1$ regarding the connections to the nodes in R_1 was obtained by collecting their best responses which was decoupled from the decision of other nodes in $R_2 \setminus R_1$ and hence is optimal under any coalition. The argument follows by induction on i to show that S cannot contain any node from R_i, $i \geq 1$, showing S must be empty.

The final remark in this section is that a mixed-strategy Nash equilibrium always exists, regardless of whether the popular-connections graph has a cycle, as shown in the next proposition.

Proposition 2. *Given any valuation matrix* **V** *and transmission parameter $\beta \in [0, 1]$, there always exists a mixed-strategy Nash equilibrium.*

The existence of a mixed-strategy Nash equilibrium follows from the game being a finite game, and then applying Kakutani's fixed point theorem.

4. Characterization of pure-strategy Nash equilibria

In this section, we address the question of how changing the transmission parameter β affects the topology of the formed network in the equilibrium. We focus on the case where $(\mathcal{V}, E_\gamma)$ does not contain any directed cycle. Using Theorem 1, this network has a pure-strategy Nash equilibrium. We show that as β increases, the network structure changes from a collection of long sparse chains to dense (and possibly smaller) components. Furthermore, we introduce two threshold functions using valuation matrix and cost parameter, and show that when transmission parameter β is lower than these thresholds, the edge set of the equilibrium network is a subset of E_γ. However, when transmission parameter β is larger than these thresholds, the equilibrium network will include low valuation edges as well, and the network is segregated into smaller dense components. To facilitate the statement of the theorem, we first introduce some notations and definitions, which we will use in the rest of the paper.

Definition 6 *(Clustering coefficient and triadic closure).* Consider a directed graph $G = (\mathcal{V}, E)$. Given the adjacency matrix $A = [a_{ij}]_{i,j \in \mathcal{V}}$ ($a_{ij} = 1$ if and only if $i \to j$), for any $i \in \mathcal{V}$, the individual clustering coefficient of i is defined as

$$C_i = \frac{\sum_{j \neq k} a_{ij} a_{jk} a_{ik}}{\sum_{j \neq k} a_{ij} a_{jk}}. \tag{2}$$

Each non-zero term of the denominator of (2) is called a *triadic*, i.e., $i \to j \to k$, is a triadic associated with agent i. If $i \to j \to k$ and $i \to k$, then we say the triadic $i \to j \to k$ is *closed* and $i \to k$ is the *closing edge of the triadic*. For instance, if all the triadics are closed, then the clustering coefficient of each agent is one.

We define the *minimum connection loss*, denoted by $\iota_{\min}$, as the minimum direct "damage" an agent incurs when establishing a low value connection, i.e.,

$$\iota_{\min} = \gamma - \max\left\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i, j \in \mathcal{V}\right\},$$

and we define the *maximum connection loss*, denoted by $\iota_{\max}$, as the maximum direct "damage" an agent incurs when establishing a low value connection, i.e.,

$$\iota_{\max} = \gamma - \min\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i, j \in \mathcal{V}\}.$$

We refer to $\beta\gamma$ as the *minimum indirect gossip cost*. We also refer to $\gamma(1 - (1 - \beta)^\mu)$ as the *maximum indirect gossip cost*, where μ is the maximum min-cut among all pairs of nodes in $(\mathcal{V}, E_\gamma)$. I.e.,

$$\mu = \max_{i,j \in \mathcal{V}} \mu(i, j),$$

where $\mu(i, j)$ is the minimum number of edges whose removal will disconnect i from j in the graph $(\mathcal{V}, E_\gamma)$.

Using these definitions, we can state the phase transition result as follows[11]:

Theorem 2 *(Phase transition in equilibrium network). Given the valuation matrix* **V***, suppose the popular-connections graph does not have any cycle. In the equilibrium network, we have:*

1. *If the* maximum indirect gossip cost *is lower than the* minimum connection loss*, i.e.,* $\gamma(1-(1-\beta)^{\mu}) < \iota_{\min}$*, then the edges of the equilibrium network is a subset of* E_{γ}*, and the individual clustering coefficient of each node is at most the individual clustering coefficient of that node in the popular-connections graph.*
2. *If the* minimum indirect gossip cost *is higher than the* maximum connection loss*, i.e.,* $\beta\gamma > \iota_{\max}$*, then in the equilibrium network the individual clustering coefficient of each node is one.*

The intuition for this theorem can be obtained by noting that the *minimum connection loss* is a lowerbound on the damage that connecting to an agent can cause (not including indirect effects), while the *maximum connection loss* is an upperbound. Therefore, when the maximum cost of indirect gossip, $\gamma(1-(1-\beta)^{\mu})$, is less than the minimum connection loss, then the equilibrium network will never contain an edge who is not present in the popular-connections graph. Conversely, when the minimum cost of indirect gossip, $\gamma\beta$, is greater than the maximum connection loss, an agent will always connect to a friend of her friend, even if she has a low direct benefit from this connection, because the cost of indirect gossip always exceeds this value, implying that it is always better to make a direct connection and obtain the benefit of this connection rather than to suffer indirect gossip. This reasoning yields the result that all triadics will be closed and thus the individual clustering coefficient of each agent will be one.

Remark 1. If $\iota_{\min} = \iota_{\max}$ and $\mu = 1$, then Theorem 2 establishes a sharp phase transition as a function of β for determination of the individual clustering coefficient of all agents.[12]

The intuition for this result is instructive. As we increase β, there are two effects loosely corresponding to the forces identified before, creating respectively strategic complementarity and substitutability. First, consider the decision of agent i to connect to agent j with whom she is indirectly connected to, i.e., there exists a directed path of friendship from i to j. As β increases, the probability that i's information will leak to j through indirect path increases, encouraging i to directly connect to j. The second force can be seen in the case when i considers connecting to individual j with whom she is not indirectly connected. In this case, by initiating a connection, i will expose herself to indirect leakage to j's friends. As β increases, this leakage becomes more likely discouraging connection from i to j.

When β is low, our result shows that the second force dominates and the network is sparse consisting only of edges in E_{γ}, i.e., agents only connect to a subset of their high valuation friends. When β is large, the first force dominates and generates a network in which clustering coefficients of all nodes are equal to one.[13]

We will demonstrate the results of Theorem 2 in the next example.

Example 4. We consider the same setting as in Example 2, where there are 9 nodes and all the high value links (shown in the column at the left hand side of Fig. 3) have valuation $H = \frac{9.5}{4}\gamma$ and all the low value links have valuation $L = \frac{1}{2}\gamma$. This choice of H and L guarantees $(H-\gamma)+2(L-\gamma) > 0$, $(H-\gamma)+3(L-\gamma) < 0$, and $2(H-\gamma)+5(L-\gamma) > 0$, which are the same assumption as in Example 2. As $\iota_{\max} = \iota_{\min} = \frac{1}{2}\gamma$, and $\mu = 2$, the thresholds on β predicted by Theorem 2 become $\beta < 1 - \frac{1}{\sqrt{2}}$ and $\beta > \frac{1}{2}$. For $\beta < 1 - \frac{1}{\sqrt{2}}$, the equilibrium network is depicted in Fig. 4a, as predicted by part (a) of Theorem 2. For any $\beta > \frac{1}{2}$, the equilibrium network is depicted in Fig. 4b, as predicted by part (b) of Theorem 2. For instance, the clustering coefficients of nodes 1 and 3 with $\beta < 1 - \frac{1}{\sqrt{2}}$ (Fig. 4a) are equal to 0, whereas the clustering coefficients of both nodes 1 and 3 with $\beta > \frac{1}{2}$ (Fig. 4b) are equal to 1.

5. Triadic closure and homophily

In this section we show the emergence of triadic closure and high clustering coefficient in equilibrium network. Note that the results of this section hold for any valuation matrix **V**, i.e., we do not impose any assumptions on the popular-connections graph.

[11] Using Proposition 1 and the construction of the Nash equilibrium given in Theorem 1, the equilibrium network characterized in Theorem 2 is also a strong Nash equilibrium.

[12] For instance, if in the graph $(\mathcal{V}, E_{\gamma})$ there is only one directed path between any two nodes, then we have $\mu = 1$.

[13] The phase transition can also be viewed as the effect of changing the cost parameter γ on the topology of network equilibrium. This is because the conditions $\gamma(1-(1-\beta)^{\mu}) < \iota_{\min}$ and $\beta\gamma > \iota_{\max}$ can be viewed as conditions on γ.

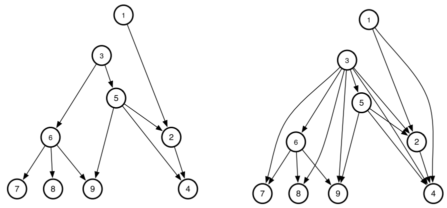


Fig. 4. Equilibrium network for the setting described in Example 4. When the transmission parameter β is sufficiently large, the clustering coefficient of all nodes is equal to one, as highlighted in Theorem 2.

5.1. Triadic closure in equilibrium network

In the next theorem, we consider triadic closure in both pure-strategy Nash equilibrium and mixed-strategy Nash equilibrium and establish sufficient conditions under which the triadics are closed in the equilibrium.

Theorem 3 *(Triadic closure in equilibrium). Given $\beta \in [0, 1]$, in any pure-strategy Nash equilibrium $\mathbf{x}$, when $x_{ij} = 1$ and $x_{jk} = 1$, then $x_{ik} = 1$ if*

$$v_{ik} \geq \gamma(1-\beta)\,(1+\beta\,\Gamma(k,\mathbf{x}))\,, \tag{3}$$

i.e., triadics are closed if the valuation of the closing edge is higher than the marginal increase of the gossip cost. Similarly, in any mixed-strategy Nash equilibrium $\boldsymbol{\sigma}$, if for $i, j, k \in \mathcal{V}$, $\mathbb{P}_\sigma[x_{ij} = 1, x_{jk} = 1] = 1$, then $\mathbb{P}_\sigma[x_{ik} = 1] = 1$, if

$$v_{ik} \geq \gamma(1-\beta)\,(1+\beta\,\mathbb{E}[\Gamma(k,\boldsymbol{\sigma})])\,.$$

The intuition for this theorem can be obtained by comparing the marginal loss and the marginal benefit i incurs by connecting to k. The marginal benefit of connecting to k for i is v_{ik}. On the other hand, once the connections $i \to j$ and $j \to k$ are present, the marginal loss of connecting to k for i can be upper bounded by $\gamma(1-\beta)\,(1+\beta\,\Gamma(k,\mathbf{x}))$. This expression can be understood by noting that the marginal loss is zero with probability β (which is the probability that i's information leaks from j to k). It is non-zero with the complementary probability $(1-\beta)$, and the loss that this will impose on i can be upper bounded by $1+\beta\,\Gamma(k,\mathbf{x})$, which captures the loss to i from his information leaking to k with probability 1 and to others through k which is at most $\beta\,\Gamma(k,\mathbf{x})$ (this is not equality because i's information may leak through other paths, not including j and k to the rest of the network).

Theorem 3 thus shows that, if condition (3) is satisfied, all triadics will be closed. This condition is in fact not very restrictive, because the term $\gamma(1-\beta)(1+\beta\,\Gamma(k,\mathbf{x}))$ will be typically small. In particular, it will tend to be small when

(a) β is close to 1, which implies that the existence of edge $i \to k$ will slightly change the cost of gossip for i, because with a high probability the information of i has already been leaked to k through the path $i \to j \to k$.
(b) $\Gamma(k,\mathbf{x})$ is small, implying that making a connection to k would leak a small amount of information. Hence, the benefit of connection to k overcomes the cost of gossip by forming a connection to k and connection happens if $v_{ik} > \gamma$.
(c) γ is small which implies that the cost of gossip is small, so that agent i would connect to agent k.

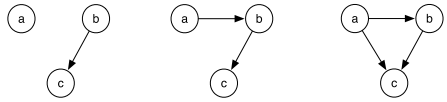
Also, note that when $\beta = 1$, the right-hand side of condition (3) will be zero, which implies that in any equilibrium triadics will be closed.

Corollary 1. *For transmission parameter $\beta = 1$, in any pure-strategy Nash equilibrium all triadics are closed, i.e., for three nodes, i, j, and k, if $x_{ij} = 1$ and $x_{jk} = 1$, then $x_{ik} = 1$. Similarly, in the mixed-strategy Nash equilibrium $\boldsymbol{\sigma}$, for $i, j, k \in \mathcal{V}$, if we have $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ij} = 1, x_{jk} = 1] = 1$, then $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ik} = 1] = 1$.*

This corollary also implies that if there is any path (not necessarily a triadic) that delivers i's information to k, then i must be connected to k. In other words, in the equilibrium if there exists a path $i = i_0 \to i_1 \to \cdots \to i_l = k$, i.e., $x_{i_j i_{j+1}} = 1$, for all $j \in \{0, \ldots, l-1\}$, then we would have $x_{ik} = 1$.

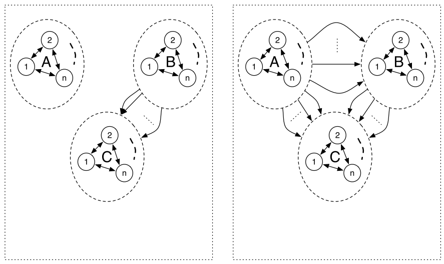
As argued before, condition (3) implies that for large β, in any equilibrium all triadics are closed. We next illustrate that even if β is small, the existence of large connected communities will induce triadic closure in any equilibrium. The reason is that if an agent connects to a member of a large connected community, then her information leaks to all members of that community with a probability close to one, and this implies that she would prefer to connect to all members of that community. This ensures that all the triadics will be closed. We will illustrate this via an example.

Example 5 *(Effect of community size).* We consider three agents $\{a, b, c\}$, where $E_\gamma = \{(a, b), (b, c)\}$ and suppose that all valuations that are greater than γ are equal to $H > \gamma$ and all valuations that are smaller than γ are equal to $0 < L < \gamma$.



(a) $H+L<2\gamma$, $H<\gamma(1+\beta)$ (b) $H+L>2\gamma$, $L<\gamma(1-\beta)$ (c) $H+L>2\gamma$, $L>\gamma(1-\beta)$

Fig. 5. Depending on the parameters, there exist three different network equilibria.



(a) $H+L<2\gamma$ (b) $H+L>2\gamma$

Fig. 6. Triadic closure among communities: for sufficiently large size of communities the individual clustering coefficient of each agent is one.

Depending on the parameters $\gamma > 0$ and $\beta > 0$, and the valuations H and L, we can have three different network equilibria as depicted in Fig. 5. In Fig. 5b, the triadic $a \to b \to c$ is not closed. We now show that if instead of three agents, we have three communities and community sizes are sufficiently large, then triadics are always closed. Suppose, in particular, that there are three communities, A, B, and C each of size n. Suppose also that valuations are given by

$$v_{ij} = \begin{cases} H & \text{if } i, j \in \text{ same group or } (i,j) \in A \times B, \text{ or } (i,j) \in B \times C, \\ L & \text{otherwise.} \end{cases}$$

In any equilibrium the agents of C are not connected to agents in other communities (as they have low valuations) and therefore they are connected to each other (as they have high valuations). Given this, in any equilibrium all agents in B will connect to each other as well as to all agents in C. We next show that in any equilibrium network for sufficiently large n all agent of A are either connected to all nodes of B and C or are not connected to any of them depending on whether $H + L - 2\gamma > 0$ or not. We consider $a \in A$ and let B_a and C_a denote the neighbors of a in B and C, respectively. We will show that if $H + L - 2\gamma > 0$, then $B_a = B$ and $C_a = C$; and if $H + L - 2\gamma < 0$, then $B_a = \emptyset$ and $C_a = \emptyset$.

- $H + L - 2\gamma > 0$: first note that if $B_a = \emptyset$, then $C_a = \emptyset$ as $v_{ac} < \gamma$ for all $c \in C$. But if $B_a = \emptyset$ and $C_a = \emptyset$, then there exists a profitable deviation for a that is to connect to all nodes in B and C to obtain the marginal utility $n(H + L - 2\gamma) > 0$. Hence $B_a \neq \emptyset$, i.e., there exist $b \in B_a$. For all $b' \in B \setminus B_a$ (if exists), we have

  $$\mathbb{P}[a \rightsquigarrow b'] \geq 1 - (1 - \beta^2)^{n-2},$$

  as there are $n-2$ distinct paths of length 3 from a to b', i.e., all paths of the form $a \to b \to b'' \to b'$ for all $b'' \in B \setminus \{b, b'\}$. For all $c' \in C \setminus C_a$ (if exists), we have

  $$\mathbb{P}[a \rightsquigarrow c'] \geq 1 - (1 - \beta^2)^{n-1},$$

  as there are $n - 1$ distinct paths of length 3 from a to c', i.e., all paths of the form $a \to b \to b'' \to c'$ for all $b'' \in B \setminus \{b\}$. Therefore, the marginal utility a receives from connecting to all nodes in B and C is lower bounded by

  $$|B \setminus B_a| \left(H - \gamma + \gamma \left(1 - (1 - \beta^2)^{n-2} \right) \right) + |C \setminus C_a| \left(L - \gamma + \gamma \left(1 - (1 - \beta^2)^{n-1} \right) \right)$$

  which is positive for sufficiently large n provided that either $C_a \neq C$ or $B_a \neq B$, showing $B_a = B$ and $C_a = C$. Finally, note that all nodes of A are connected to each other since it does not incur any additional gossip cost and the equilibrium network is the one shown in Fig. 6b.
- $H + L - 2\gamma < 0$: similar to the previous part, for $a \in A$ if there exists $b \in B_a$, then for sufficiently large n we would have $B_a = B$ and $C_a = C$. However, a profitable deviation in this case is to disconnect from all nodes in B and C to obtain 0 instead of $n(H + L - 2\gamma) < 0$, contradicting the assumption that there exists $b \in B$. Therefore, the only remaining case is to have $B_a = \emptyset$ and consequently $C_a = \emptyset$. Again, note that all nodes of A are connected to each other since it does not incur any additional gossip cost and the equilibrium network is the one shown in Fig. 6a.

5.2. Homophily

As noted above, homophily refers to a situation in which agents are more likely to be friends with or have links with others in their community than those outside their community. In this subsection, we show that when there is a slight

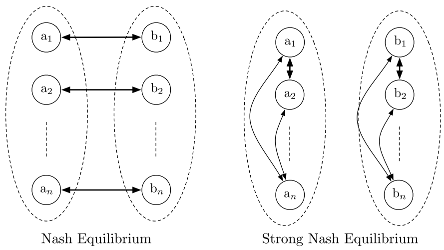


Fig. 7. Nash equilibrium vs. strong Nash equilibrium.

difference in terms of direct benefits from connecting within the community, this will lead to a significant pattern of homophily because of privacy concerns.

To establish these results, we will focus on the strong Nash equilibrium of the network formation game because, as the next example illustrates, there will sometimes also exist other, unintuitive pure-strategy Nash equilibria, but these are never strong Nash equilibria.

Example 6 *(Nash equilibrium versus strong Nash equilibrium).* Consider two groups of agents each of them of size $n > 1$, denoted by $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$. We let valuations within each group be H and valuations across groups be L (symmetric valuations), except that valuation between a_i and b_i are H for $i = 1, \dots, n$. Also, let $0 < L < \gamma < H$, $\beta = 1$, and $2H + L - 3\gamma < 0$. One pure-strategy Nash equilibrium, for this network is shown in Fig. 7 where a_i is connected to b_i for all $i = 1, \dots, n$ and there is no other connection between agents. However, it is not a strong Nash equilibrium, because if all agents $a_1, \dots, a_n$ deviate together, disconnect from B and connect to each other, then the utility of each of them would be improved to $n(H - \gamma)$, instead of $H - \gamma$ in the Nash equilibrium shown in Fig. 7. The strong Nash equilibrium network for this setting is two segregated completely connected communities as depicted in Fig. 7.

Intuitively, in strong Nash equilibrium, privacy concerns force the society to form two separate communities. The reason is that if an agent such as a is part of a community A and the rest of community A is connected to another community B, then a would benefit from connecting to all individuals in B as well, as her information is already leaking to all agents in B through her connections with people of her own community. However, if nodes in A decide altogether to disconnect from nodes in B, then this would be beneficial for all of them. This example thus illustrates why the notion of strong Nash equilibrium plays an important, albeit intuitive, role in the emergence of clustered networks and homophily.

Example 6 motivates our focus on strong Nash equilibria in the rest of this subsection. We will then see that strong Nash equilibria will feature a strong form of homophily.

Let us now focus on the two-community society (with the two groups denoted by A and B) with the following probabilistic valuation pattern:

$$v_{ij} = \begin{cases} L & \text{w.p. } p & \text{if } i, j \in \text{ same group} \\ H & \text{w.p. } 1-p & \text{if } i, j \in \text{ same group} \\ L & \text{w.p. } 1-p & \text{if } i, j \in \text{ different groups} \\ H & \text{w.p. } p & \text{if } i, j \in \text{ different groups}, \end{cases}$$

where L corresponds to low valuation, and $H > L$ to high valuation, and $p \in [0, 1]$. These preferences thus indicate that there is some "homophily" in preferences, but this is quite weak because H could be arbitrarily close to L. More specifically, within each community, agents have on average pn low valuations and across the communities agents have on average pn high valuations. However, we show that even in this setting the equilibrium network will be highly clustered and will feature homophily. We should emphasize that, though we are considering a probabilistic setting in terms of the valuation matrix, the game is still one of complete information, i.e., the players know the $\mathbf{V}$ matrix when making their choices.

Theorem 4. *Consider the probability distribution of v_{ij}'s described above. For any $\eta > 0$, there exists n_0 such that for any $n \geq n_0$ with probability at least $1 - \eta$ we have.*

(a) If $H, L > \gamma$, then a complete network is the only strong Nash equilibrium.

(b) If $L < \gamma < H$, then for $p < \frac{\min\{\beta\gamma, \gamma - L\}}{2(H-\gamma) + \left(3 + \frac{H-L}{H-\gamma}\right)\min\{\beta\gamma, \gamma - L\}}$, in any strong Nash equilibrium there exist $S_A \subseteq A$ and $S_B \subseteq B$, each with size at least $n\left(1 - p\left(\frac{H-\gamma}{\min\{\beta\gamma, \gamma - L\}} + 2 + \frac{H-L}{H-\gamma}\right)\right)$ such that there exist no connection from A to S_B and no connection from B to S_A.

The intuition is as follows: because of the slight preference for within-community links, there will be more within-community connections, but this in turn implies that an agent will have further incentives to form within-community links because his/her information is already likely to have been leaked to potential friends within the community. In contrast, he/she will refrain from links to the other community, because even a single across-community link will imply the leakage of his/her information to many other connected people within this other community.

Note also that the parameter p needs to be smaller than a certain threshold for this result to hold. Clearly, p needs to be less than 1/2, but the theorem specifies a lower threshold, which helps to ensure that there are sufficiently more within-community links than cross-community links in any (strong) Nash equilibrium.

6. Conclusion

The technological developments of the last decade and a half have increased the sharing of information over various social media. With this trend set to continue, concerns over privacy have also mounted, and are expected to become a growing constraint on the functioning of many online platforms. Despite the centrality of issues of privacy both in online platforms and in various real-world and virtual social networks, there is relatively little game-theoretic analysis of privacy and efforts by agents to protect their privacy.

In this paper, we took the first step in this direction by modeling how privacy concerns affect individual choices in the context of a network formation game (where links can be interpreted as friendships in a social network, connections over a social media platform or trading activities in an online platform). In the model, each individual decides which other agents to "befriend", i.e., form links with. Such links bring direct (heterogeneous) benefits from friendship and also lead to the sharing of information. But such information can travel over other linkages (e.g., shared by the party acquiring the information with others), defining a percolation process over the equilibrium network. Privacy concerns are modeled as a disutility that an individual suffers as a result of her private information being acquired by others, and imply that the individual has to take into account who the friends of her new friend (and who the friends of friends of her new friend etc.) are.

After showing that pure-strategy Nash equilibria may fail to exist, we provided sufficient (and necessary) conditions for the existence of pure-strategy equilibria, and characterized their structure. Information flows over the social networks create both strategic complementarities and substitutabilities, which in turn lead to a phase transition result whereby small changes in the transmission of information over the network can fundamentally change the nature (and clustering) of the equilibrium network.

Our main results concern the analysis of triadic closure and homophily. We show that clustering of links and thus triadic closure emerges naturally because if player a is friend with b and b is friend with c, then a's information is likely to be shared indirectly with c anyway, thus making it less costly for a to befriend c. Homophily also emerges as part of the equilibrium network formation (provided that we focus on strong Nash to avoid other potential equilibria with the flavor of coordination failure). This is because even an infinitesimal advantage in terms of direct benefits of friendship within a group makes linkages within that group more likely, in turn making information travel within that group, reducing the cost of making further within-group links due to loss of privacy, and thus increasing the likelihood of further within-group links.

Appendix A. Proofs of Section 3

Proof of Theorem 1. We first show a simple lemma which we will use in the proof.

Lemma 1. *Let $G = (\mathcal{V}, E)$ be a directed graph. For any $a, b \in \mathcal{V}$, $\mathbb{P}[a \rightsquigarrow b]$ increases (does not decrease) if we add any edge such as $c \to d$.*

Proof. We let $P_a^{E_\mathbf{x}}(E)$ denote the probability of the realized graph being $G = (\mathcal{V}, E)$ when the strategy profile is $E_\mathbf{x}$ (note that this probability depends on node a, hence the notation). For a given strategy profile $E_\mathbf{x}$, where $(c, d) \notin E_\mathbf{x}$, we have

$$\begin{aligned}
&\sum_{E \subseteq E_\mathbf{x}} P_a^{E_\mathbf{x}}(E)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} \\
&= \sum_{E \subseteq E_\mathbf{x}} P_a^{E_\mathbf{x}}(E)\beta\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} + \sum_{E \subseteq E_\mathbf{x}} P_a^{E_\mathbf{x}}(E)(1-\beta)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} \\
&\le \sum_{E \subseteq E_\mathbf{x}} P_a^{E_\mathbf{x}}(E)\beta\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E \cup \{(c,d)\})\} + \sum_{E \subseteq E_\mathbf{x}} P_a^{E_\mathbf{x}}(E)(1-\beta)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} \\
&= \sum_{E \subseteq E_\mathbf{x} \cup \{(c,d)\},\ (c,d) \in E} P_a^{E_\mathbf{x} \cup \{(c,d)\}}(E)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} + \sum_{E \subseteq E_\mathbf{x} \cup \{(c,d)\},\ (c,d) \notin E} P_a^{E_\mathbf{x} \cup \{(c,d)\}}(E)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\} \\
&= \sum_{E \subseteq E_\mathbf{x} \cup \{(c,d)\}} P_a^{E \cup \{(c,d)\}}(E)\mathbf{1}\{a \rightsquigarrow b \text{ in } G = (\mathcal{V}, E)\},
\end{aligned}$$

which completes the proof of lemma. □

Proof of part 1. We first prove the necessary condition by showing that if there exists a directed cycle of length at least three in popular-connections graph, then there exists a set of valuations $\mathbf{V}$ and transmission probability β for which pure-strategy Nash equilibrium does not exist. Given popular-connections graph, consider the smallest cycle in this graph, denoted as $C = (\mathcal{V}_c, E_c)$, where $\mathcal{V}_c = \{1, \ldots, k\}$, where k is the size of this cycle (without loss of generality we let $\mathcal{V}_c = \{1, \ldots, k\}$). We now assign v_{ij}'s that are consistent with $(\mathcal{V}, E_\gamma)$.

1. For all edges that do not belong to E_γ let $v_{ij} = 0$. As we have $\gamma > 0$, this guarantees that in any Nash equilibrium, there exists no edge between nodes i and j where $v_{ij} = 0$.
2. For $(i, j) \in E_\gamma$, where either $i \notin \mathcal{V}_c$ or $j \notin \mathcal{V}_c$, let $v_{ij} = \infty$ (i.e., large enough). This guarantees that in any Nash equilibrium $x_{ij} = 1$.

Next, we show that there exists an assignment of v_{ij}'s for $i, j \in \mathcal{V}_c$ for which no pure-strategy Nash equilibrium exist.

First note that as $(\mathcal{V}_c, E_c)$ is the smallest cycle in $(\mathcal{V}, E_\gamma)$, for all $i, j \in \mathcal{V}_c$ such that $i \neq j+1 \pmod k$, $(i, j) \notin E_\gamma$ (otherwise we can find a cycle smaller than C which contradicts the fact that C is the smallest cycle). Therefore, we have assigned $v_{ij} = 0$ to those pairs, implying there is no connection among $i, j \in \mathcal{V}_c$ such that $i \neq j+1 \pmod k$. We let $v_{i,i+1} \geq \gamma$ for $i = 1, \ldots, k$ (all i's are modulus k). We now specify these valuations. We consider two cases, depending on whether k is odd or even.

(i) k is an odd number: let $\mathbf{x}$ show the connections of agents without considering connections among nodes in $\mathcal{V}_c$.

Claim. *There exists $\bar{\beta} > 0$ such that for all $\beta < \bar{\beta}$ and $i = 1, \ldots, k$, we have*

$$\begin{aligned}&\min_{\mathbf{x}^c} \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 1, \mathbf{x}^c) \\ &> \max_{\mathbf{x}^c} \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 0, \mathbf{x}^c),\end{aligned} \tag{A.1}$$

where $\mathbf{x}^c = (x_{j,(j+1)} \ : \ j \notin \{i, i+1\})$ shows all the connections except the ones involving i and $i+1$.

Proof of Claim. first note that using Lemma 1 both sides are non-negative. We now evaluate the left hand side of (A.1). We let S_i denote the set of out-neighbors of i for strategy vector $\mathbf{x}$, when $x_{i,i+1} = 0$. Since C is the smallest cycle in $(\mathcal{V}, E_\gamma)$, we have $\mathcal{V}_c \cap S_i = \emptyset$, showing S_i does not depend on the choice of $\mathbf{x}^c$. Also, S_i does not depend on $x_{i+1,i+2}$. For β sufficiently small, we have

$$\begin{aligned}\Gamma(i, \mathbf{x}, x_{i,i+1} = 1, \mathbf{x}^c \cup x_{i+1,i+2}) &= |S_i| + 1 + \beta \sum_{k \in \mathcal{V}} \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{i,k} x_{k,j} + O(\beta^2) \\ &= |S_i| + 1 + \beta \sum_{k \in S_i} \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{k,j} + \beta \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{i+1,j} + O(\beta^2).\end{aligned}$$

Similarly, when $x_{i,i+1} = 0$, we have

$$\Gamma(i, \mathbf{x}, x_{i,i+1} = 0, \mathbf{x}^c \cup x_{i+1,i+2}) = |S_i| + \beta \sum_{k \in \mathcal{V}} \sum_{j \in \mathcal{V} \setminus S_i} x_{i,k} x_{k,j} + O(\beta^2) = |S_i| + \beta \sum_{k \in S_i} \sum_{j \in \mathcal{V} \setminus S_i} x_{k,j} + O(\beta^2).$$

Therefore, for any given $\mathbf{x}^c$ we have

$$\begin{aligned}&\Gamma(i, \mathbf{x}, x_{i,i+1} = 1, \mathbf{x}^c \cup x_{i+1,i+2}) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, \mathbf{x}^c \cup x_{i+1,i+2}) \\ &= |S_i| + 1 + \beta \sum_{k \in S_i} \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{k,j} + \beta \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{i+1,j} - \left(|S_i| + \beta \sum_{k \in S_i} \sum_{j \in \mathcal{V} \setminus S_i} x_{k,j} \right) + O(\beta^2) \\ &= 1 - \beta \sum_{k \in S_i} x_{k,i+1} + \beta \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1\})} x_{i+1,j} + O(\beta^2).\end{aligned}$$

Note that S_i only depends on $\mathbf{x}$ and is independent of $\mathbf{x}^c$ and $x_{i+1,i+2}$. For any choice of $\mathbf{x}^c$, we then have

$$\begin{aligned}&\min_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 1, \mathbf{x}^c) \right\} \\ &= 1 - \beta \sum_{k \in S_i} x_{k,i+1} + \beta + \beta \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1,i+2\})} x_{i+1,j} + O(\beta^2) \\ &\geq 1 - \beta \sum_{k \in S_i} x_{k,i+1} + \beta \sum_{j \in \mathcal{V} \setminus (S_i \cup \{i+1,i+2\})} x_{i+1,j} + O(\beta^2) \\ &= \max_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 0, \mathbf{x}^c) \right\},\end{aligned}$$

completing the proof of claim. □

We now proceed with the proof of Theorem. We let

$$\Delta^{\min}_{i,i+1} = \min_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 1, \mathbf{x}^c) \right\},$$

and

$$\Delta^{\max}_{i,i+1} = \max_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{i,i+1} = 0, x_{i+1,i+2} = 0, \mathbf{x}^c) \right\}.$$

For any $i = 1, \dots, k$, we also let

$$v_{i,i+1} \in \left(\gamma \Delta^{\max}_{i,i+1}, \gamma \Delta^{\min}_{i,i+1} \right).$$

We now completed matrix **V** which is compatible with popular-connections graph. We next show that there exists no pure-strategy Nash equilibrium for valuation matrix **V**. First, note that the network with no edge from the cycle is not an equilibrium. This is because any agent i benefits from deviating by connecting to $i+1$, and obtaining extra utility of at least $v_{i,i+1} - \gamma \Delta^{\max}_{i,i+1} > 0$. Therefore, we can find i such that $i \to i+1$. Without loss of generality, suppose $k \to 1$ and let $\mathbf{x}^*$ denote the equilibrium decisions. This would imply that $k-1 \nrightarrow k$, because agent $k-1$ faces a loss of at least $\gamma \Delta^{\min}_{k-1,k}$, compared to the benefit of $v_{k-1,k}$. I.e., if $k-1$ connects to k, his/her utility would increase by at most $v_{k-1,k} - \gamma \Delta^{\min}_{k-1,k}$, which is negative. Note that the only possible deviation for node $k-1$ is regarding connecting to node k. This is because by the choice of matrix **V** decision regarding connecting to any other node (irrelevant of the others' decisions) is predetermined.

This in turn shows the following set of decisions $k-2 \to k-1, k-3 \nrightarrow k-2, \dots, 1 \to 2, k \nrightarrow 1$, which contradicts the assumption $k \to 1$, showing that no pure-strategy Nash equilibrium exists for the assigned valuations and $\beta \le \bar{\beta}$.

(ii) If k is even, similar to the proof of claim, we can show there exists $\bar{\beta}$ such that for $\beta < \bar{\beta}$ we have

$$\min_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 0, x_{i+1,i+2} = 1, \mathbf{x}^c) \right\}$$
$$> \max_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 0, x_{i+1,i+2} = 0, \mathbf{x}^c) \right\},$$

and

$$\min_{\mathbf{x}^c} \left\{ \Gamma(k, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 1, x_{2,3} = 1, \mathbf{x}^c) - \Gamma(k, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 0, x_{2,3} = 1, \mathbf{x}^c) \right\}$$
$$> \max_{\mathbf{x}^c} \left\{ \Gamma(k, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 1, x_{2,3} = 0, \mathbf{x}^c) - \Gamma(k, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 0, x_{2,3} = 0, \mathbf{x}^c) \right\}.$$

We let $v_{1,2} = \infty$, i.e., large enough so that we always have $x_{12} = 1$. For ease of notation, for $i \neq k$, we define

$$\Delta^{\min}_{i,i+1} = \min_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 0, x_{i+1,i+2} = 1, \mathbf{x}^c) \right\},$$
$$\Delta^{\max}_{i,i+1} = \max_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 1, x_{i+1,i+2} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{i,i+1} = 0, x_{i+1,i+2} = 0, \mathbf{x}^c) \right\}.$$

Also, for $i = k$ we define

$$\Delta^{\min}_{k,1} = \min_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 1, x_{2,3} = 1, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 0, x_{2,3} = 1, \mathbf{x}^c) \right\},$$
$$\Delta^{\max}_{k,1} = \max_{\mathbf{x}^c} \left\{ \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 1, x_{2,3} = 0, \mathbf{x}^c) - \Gamma(i, \mathbf{x}, x_{1,2} = 1, x_{k,1} = 0, x_{2,3} = 0, \mathbf{x}^c) \right\}.$$

Now we let $v_{1,2} = \infty$ (i.e., large enough to ensure we always have $x_{12} = 1$) and for $i = 2, \dots, k-1$ we let

$$v_{i,i+1} \in (\gamma \Delta^{\max}_{i,i+1}, \gamma \Delta^{\min}_{i,i+1}).$$

We now consider two possible cases and show that in both of them equilibrium does not exist.

- Either $2 \to 3$: This would imply that $k \nrightarrow 1$, which in turn shows that $k-1 \to k, \dots, 3 \to 4, 2 \nrightarrow 3$, which is a contradiction,
- Or $2 \nrightarrow 3$: This would imply that $k \to 1$, which in turn shows that $k-1 \nrightarrow k, \dots, 3 \nrightarrow 4, 2 \to 3$, which is again a contradiction, completing the proof.

Therefore, we showed that if there exists a cycle in popular-connections, we can construct **V** compatible with popular-connections, for which pure-strategy Nash equilibrium does not exist.

Proof of part 2. We next show that the absence of directed cycles in popular-connections is sufficient for the existence of pure-strategy Nash equilibrium. We will construct an update rule that converges to an equilibrium in finitely many steps. The construction is as follows.

Because the popular-connections graph has no directed cycle, there exist some nodes with no outgoing edges (otherwise, we would have a cycle). Note that if there are multiple components in $(\mathcal{V}, E_\gamma)$ that are not connected, then we will perform the construction on each of them separately. We let R_1 denote the set of nodes with no outgoing edges. We also let $H_1 = (\mathcal{V}, E_\gamma)$. In each step of the construction, we update the set R and the graph H. In the first step of the construction, we let all nodes to play their optimal decisions regarding the nodes in R_1. We now update H_1 by adding the newly created edges to obtain H_2. We then update R_1 to be the set of nodes with no outgoing edges plus the nodes with outgoing edges only in the set R_1, and denote the updated set by R_2. We now proceed to the next step of construction. In the second step, we let all nodes to play their optimal decisions regarding the nodes in R_2. We then update H_2 by adding the newly created edges. Finally, we update R_2 to be the set of nodes with outgoing edges only in the set R_2 as well as the nodes with no outgoing edges. We continue the steps until R becomes the entire set of nodes $\mathcal{V}$. We claim that the resulting graph is an equilibrium network.

We first show that the construction of the sets R_i's is well-defined, meaning that at each step i we have either reached the entire set of nodes or there exists a node in $\mathcal{V} \setminus R_i$ whose all out-neighbors are in R_i. This follows from the graph not having a directed cycle. Because if for all nodes in $\mathcal{V} \setminus R_i$ there exists at least one out-neighbor in $\mathcal{V} \setminus R_i$, then there exist a directed cycle among some nodes belonging to $\mathcal{V} \setminus R_i$ which contradicts the fact that the graph $(\mathcal{V}, E_\gamma)$ has no directed cycle.

We next show that a node a that is about to be added to R at step r, will only (potentially) connect to nodes in the set R_r. We will show this by induction on r. For $r = 1$ it is evident. Suppose it holds for all $i \leq r - 1$. We will show that it holds for $i = r$ as well. As node a does not have any high value friend in $\mathcal{V} \setminus R_r$ and none of the nodes in R_r are connected to nodes in $\mathcal{V} \setminus R_r$ (by induction hypothesis), node a will only make connections to nodes in R_r. Therefore, it suffices to show that a will not deviate from the optimal decision he/she makes at step r. This holds because a is playing his/her best response regarding connecting to the nodes in R_r and all the nodes in R_r will not change their connections in the subsequent steps.

Next, we show that the best response dynamics converges to an equilibrium. In the first round of best response dynamics, all the nodes will play their optimal decisions regarding connecting to the nodes in set R_1, same as the first step of the construction, and do not change their decisions in the rest of the best response dynamics. Because nodes do not change their decisions regarding connecting to nodes in R_1, in the second round, all of the nodes will play their best response regarding connecting to nodes in R_2 and do not change their decisions in the next rounds. By repeating this argument, after finite number of rounds the created edges would be the same as the ones that the construction steps described before, would create. This shows that the resulting network from following best response dynamics in finitely many rounds becomes a pure-strategy Nash equilibrium. □

Proof of Proposition 1. We use the same notation as in the proof of Theorem 1. Consider the constructed Nash equilibrium in the proof of part 2 of Theorem 1. We will show that it is a strong Nash equilibrium. Consider a set of agents $S \subseteq \mathcal{V}$. We will show that the group S of agents does not have a profitable deviation. Note that no matter what the decisions of other nodes are, the nodes in $S \cap R_1$ will not change their decisions. Now that these nodes do not change their decisions, all nodes in $S \cap R_2$ will not change their decisions. By repeating this argument, none of the nodes in $\cup_r (S \cap R_r) = S$ (as $\cup_r R_r = \mathcal{V}$) will change their decision, showing that the resulting equilibrium in Theorem 1 is a strong Nash equilibrium. □

Appendix B. Proofs of Section 4

Proof of Theorem 2. We first introduce two notations and then proceed with proving each part of theorem separately. We let

$$\ell_{\max} = \max\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i, j \in \mathcal{V}\},$$

and

$$\ell_{\min} = \min\{v_{ij} \mid v_{ij} < \gamma,\ i \neq j,\ i, j \in \mathcal{V}\}.$$

(1) We consider the construction process described in Theorem 1 and show that if the maximum indirect gossip cost is lower than the minimum connection loss, i.e., $\gamma(1 - (1-\beta)^\mu) < \iota_{\min}$, no edge from $\mathcal{V} \times \mathcal{V} \setminus E_\gamma$ will be included in the equilibrium. Note that if the popular-connection graph does not contain any directed cycle, then the equilibrium network is unique. Hence, without loss of generality, we can follow the procedure described in Theorem 1. We use the same notation as the one used in the proof of Theorem 1 given in Appendix A. By induction on the step number, we will show that for any $r \geq 1$ at step r no node will connect to a node in $\mathcal{V} \times \mathcal{V} \setminus E_\gamma$. This evidently holds for $r = 1$. Suppose this holds for all steps $r < i$. We will show that it holds for step $r = i$ as well. Suppose that node a in step i wants to connect to node $b \in R_i$ that is not a neighbor of a in $(\mathcal{V}, E_\gamma)$. Also, suppose b is the highest ranked node added to R with this property. We have $v_{ab} < \gamma$. If a connects to b, then we have

$$\Gamma(a, \mathbf{x}, x_{ab} = 1) - \Gamma(a, \mathbf{x}, x_{ab} = 0) \geq (1-\beta)^\mu,$$

where μ is the maximum min-cut among all pairs of nodes in $(\mathcal{V}, E_\gamma)$. The reason is based on the following three facts:

1) $\mathbb{P}[a \rightsquigarrow b|\mathbf{x}, x_{ab}=1]=1$ as there is a direct link from a to b.
2) $\mathbb{P}[a \rightsquigarrow b|\mathbf{x}, x_{ab}=0] \le 1-(1-\beta)^{\mu(a,b)}$ by the induction assertion that there exist no edges besides the ones in E_γ for all steps $r<i$ and b is the highest ranked node added to set R. Note that given $x_{ab}=0$, $a \not\rightsquigarrow b$ if all edges of a min-cut between a and b in $(\mathcal{V}, E_\gamma)$ are not active. Since the size of a min-cut between a and b is at most μ (i.e., $\mu(a,b) \le \mu$) and each edge is not active with probability $1-\beta$, we have

$$\mathbb{P}[a \not\rightsquigarrow b|\mathbf{x}, x_{ab}=0] \ge (1-\beta)^{\mu}.$$

3) $\mathbb{P}[a \rightsquigarrow c|\mathbf{x}, x_{ab}=1] \ge \mathbb{P}[a \rightsquigarrow c|\mathbf{x}, x_{ab}=0]$, for any $c \neq b$ as more connections make the gossip more probable (Lemma 1).

Therefore, the marginal utility of a from connecting to b is

$$\begin{aligned} u_a(\mathbf{x}, x_{ab}=1) - u_a(\mathbf{x}, x_{ab}=0) &= v_{ab} - \gamma\left(\Gamma(a,\mathbf{x},x_{ab}=1) - \Gamma(a,\mathbf{x},x_{ab}=0)\right) \\ &\le v_{ab} - \gamma(1-\beta)^{\mu} \le \ell_{\max} - \gamma(1-\beta)^{\mu} = \gamma - \iota_{\min} - \gamma(1-\beta)^{\mu} < 0, \end{aligned}$$

as we have $\gamma\left(1-(1-\beta)^{\mu}\right) < \iota_{\min}$. This shows that the resulting Nash equilibrium only contains a subset of the edges in E_γ.

(2) We show that by following the construction given in the proof of Theorem 1, all triadics are closed. We show by induction that by following the procedure given in the proof of Theorem 1 at step r, all nodes must have closed all the possible triadics, for any r. This evidently hold for $r=1$. Suppose that this statement holds for $r \le i$, we will show that it holds for $r=i+1$. Suppose at step r, node a is connected to $b \in R_r$, b is connected to $c \in R_r$, and $a \not\to c$. Also, suppose that c is the first node (lowest ranked) added to the set R with this property. Because $a \not\to c$ and $a \to b \to c$, we have that $\mathbb{P}[a \rightsquigarrow c] \ge \beta$. Next, we show that a can make a profitable deviation by connecting to c. If a connects to c, then $\mathbb{P}[a \rightsquigarrow c]=1$ and a receives a direct benefit equal to v_{ac} from connecting to c. Moreover, because c is the first node with this property, we have $\mathbb{P}[a \rightsquigarrow d|x_{ac}=1] = \mathbb{P}[a \rightsquigarrow d|x_{ac}=0]$, for all d. Therefore, the marginal utility of a from connecting to c is

$$\begin{aligned} u_a(\mathbf{x}, x_{ac}=1) - u_a(\mathbf{x}, x_{ac}=0) &= v_{ac} - \gamma\left(\Gamma(a,\mathbf{x},x_{ac}=1) - \Gamma(a,\mathbf{x},x_{ac}=0)\right) \\ &\ge \ell_{\min} - \gamma(1-\beta) = \gamma - \iota_{\max} - \gamma(1-\beta) > 0, \end{aligned}$$

as we have $\iota_{\max} < \gamma\beta$. This shows that a would connect to c as well. Therefore, all triadics are closed in the equilibrium network.

This completes the proof. □

Appendix C. Proofs of Section 5

Proof of Proposition 3. We first show a simple lemma that we will use in the proof.

Lemma 2. *Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a pure-strategy Nash equilibrium for a given set of v_{ij}'s and a given β. For any i and j, we have $x_{ij}=1$ if*

$$v_{ij} - \gamma\left(\Gamma(i,\mathbf{x},x_{ij}=1) - \Gamma(i,\mathbf{x},x_{ij}=0)\right) \ge 0.$$

Proof. We have $x_{ij}=1$ if

$$u_i(\mathbf{x}, x_{ij}=1) - u_i(\mathbf{x}, x_{ij}=0) \ge 0.$$

Therefore, $x_{ij}=1$ if $v_{ij} - \gamma\left(\Gamma(i,\mathbf{x},x_{ij}=1) - \Gamma(i,\mathbf{x},x_{ij}=0)\right) \ge 0$. This completes the proof of lemma. □

We now proceed with the proof of Theorem.

Proof of part (a). Using Lemma 2, $x_{ik}=1$ if

$$v_{ik} \ge \gamma\left(\Gamma(i,\mathbf{x},x_{ik}=1,x_{ij}=1,x_{jk}=1) - \Gamma(i,\mathbf{x},x_{ik}=0,x_{ij}=1,x_{jk}=1)\right).$$

We also have that

$$\begin{aligned} &\Gamma(i,\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=1) - \Gamma(i,\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=0) \\ &= \sum_{l \in \mathcal{V}} \mathbb{P}[i \rightsquigarrow l|\mathbf{x}, x_{ij}=1, x_{jk}=1, x_{ik}=1] - \mathbb{P}[i \rightsquigarrow l|\mathbf{x}, x_{ij}=1, x_{jk}=1, x_{ik}=0] \\ &= \left(\mathbb{P}[i \rightsquigarrow k|\mathbf{x}, x_{ij}=1, x_{jk}=1, x_{ik}=1] - \mathbb{P}[i \rightsquigarrow k|\mathbf{x}, x_{ij}=1, x_{jk}=1, x_{ik}=0]\right) \end{aligned}$$

$$+\sum_{\substack{l\in\mathcal{V}\\ l\neq k}}\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=1]-\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=0]$$

$$\le(1-\beta)+\sum_{\substack{l\in\mathcal{V}\\ l\neq k}}\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=1]-\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=0]$$

$$\le(1-\beta)+(1-\beta)\beta\Gamma(k,\mathbf{x}),$$

where the first inequality follows from $\mathbb{P}[i\rightsquigarrow k|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=1]=1$ and $\mathbb{P}[i\rightsquigarrow k|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=0]\ge\beta$. The second inequality holds because with probability β, $j\to k$ which makes the two terms $\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=1,j\to k]$ and $\mathbb{P}[i\rightsquigarrow l|\mathbf{x},x_{ij}=1,x_{jk}=1,x_{ik}=0,j\to k]$ equal. With probability $1-\beta$, $j\not\to k$ and the difference between the two terms is bounded by the gossip of k with the extra factor of β to account for the first hop going out from k. Therefore, if $v_{ik}\ge\gamma(1-\beta)(1+\beta\Gamma(k,\mathbf{x}))$, we have $x_{ik}=1$.

Proof of part (b). Letting $\mathbf{x}_i^*\in\text{support}(\sigma_i)$, we have

$$\mathbf{x}_i^*\in\operatorname*{argmax}_{\mathbf{x}_i\in\{0,1\}^n}u_i(\mathbf{x}_i,\boldsymbol{\sigma}_{-i}),$$

where

$$u_i(\mathbf{x}_i,\boldsymbol{\sigma}_{-i})=\mathbb{E}_{\boldsymbol{\sigma}}[u_i(\mathbf{x}_i,\mathbf{x}_{-i})]=\sum_{j\neq i}v_{ij}x_{ij}-\gamma\mathbb{E}_{\boldsymbol{\sigma}}[\Gamma(i,\mathbf{x})].$$

Using Lemma 2, we have $x_{ik}^*=1$ if

$$v_{ik}-\gamma\mathbb{E}_{\boldsymbol{\sigma}}[\Gamma(i,\mathbf{x},x_{ik}^*=1)-\Gamma(i,\mathbf{x},x_{ik}^*=0)]\ge0.$$

Because $\mathbb{P}_\sigma[x_{ij}=1,x_{jk}=1]=1$, using the same argument as in part (a), we obtain

$$\mathbb{E}_{\boldsymbol{\sigma}}[\Gamma(i,\mathbf{x},x_{ik}^*=1)-\Gamma(i,\mathbf{x},x_{ik}^*=0)]\le(1-\beta)+(1-\beta)\beta\mathbb{E}_{\boldsymbol{\sigma}}[\Gamma(k,\boldsymbol{\sigma})].$$

Therefore, if $v_{ik}\ge\gamma(1-\beta)(1+\beta\mathbb{E}_{\boldsymbol{\sigma}}[\Gamma(k,\boldsymbol{\sigma})])$, then for any $\mathbf{x}_i^*\in\text{support}(\sigma_i)$, we have $x_{ik}^*=1$, which in turn shows $\mathbb{P}_{\boldsymbol{\sigma}}[x_{ik}=1]=1$. □

Proof of Theorem 4.

(a) Because both L and H are greater than γ, $x_{ij}=1$ for all $i,j\in\mathcal{V}$ is the only Nash equilibrium.
(b) We use the following Chernoff–Hoeffding bound in this proof (see e.g. Dembo and Zeitouni, 1998).

Lemma 3 *(Dembo and Zeitouni, 1998). Let $Z_1,\dots,Z_n$ be independent Bernoulli ($Z_i\in\{0,1\}$) random variables with $\mathbb{P}[Z_i=1]=p_i$. Then for any $0<\delta<1$, we have that*

$$\mathbb{P}\left[\sum_{i=1}^n Z_i\le(1-\delta)\sum_{i=1}^n p_i\right]\le\exp\left(-\frac{\delta^2}{3}\sum_{i=1}^n p_i\right),$$
$$\mathbb{P}\left[\sum_{i=1}^n Z_i\ge(1+\delta)\sum_{i=1}^n p_i\right]\le\exp\left(-\frac{\delta^2}{3}\sum_{i=1}^n p_i\right).$$

Given this lemma, we now prove the theorem. For any $a\in A$, we let $r_a^H(A)$ denote the ratio of H's among all $v_{aa'}$ for $a'\in A$. We also let $r_a^L(B)$ denote the ratio of L's among all v_{ab} for $b\in B$. Similarly, we define $r_b^H(B)$ and $r_b^L(A)$ and let $\delta>0$ to be a small enough number (we will specify later). For any $a\in A$, we let $Z_a(A)=\mathbf{1}\{r_a^H(A)\ge(1-p-\delta)\}$ which are independent binary random variables. Similarly, we define $Z_b(B)=\mathbf{1}\{r_b^H(B)\ge(1-p-\delta)\}$. Also define $Z_a(B)=\mathbf{1}\{r_a^L(B)\ge(1-p-\delta)\}$, and $Z_b(A)=\mathbf{1}\{r_b^L(A)\ge(1-p-\delta)\}$. Using Lemma 3, for any a, we have

$$\mathbb{P}[Z_a(A)=0]=\mathbb{P}\left[r_a^H(A)\le(1-p-\delta)\right]=\mathbb{P}\left[\sum_{a'\in A}\mathbf{1}\{v_{aa'}=H\}\le(n-1)(1-p-\delta)\right]$$
$$\le\exp\left(-(n-1)\frac{\delta^2}{3(1-p)}\right).\tag{C.1}$$

Similarly, for any a, we have

$$\mathbb{P}[Z_a(B)=0] = \mathbb{P}\left[r_a^L(B) \le (1-p-\delta)\right] = \mathbb{P}\left[\sum_{b\in B} \mathbf{1}\{v_{ab}=L\} \le n(1-p-\delta)\right] \le \exp\left(-n\frac{\delta^2}{3(1-p)}\right). \tag{C.2}$$

We next define a joint event as

$$\mathcal{E} = \bigcap_{a\in A}\{Z_a(A)=1\}\bigcap_{b\in B}\{Z_b(B)=1\}\bigcap_{a\in A}\{Z_a(B)=1\}\bigcap_{b\in B}\{Z_b(B)=1\}.$$

Using union bound along with eq. (C.1) and (C.2), we obtain

$$\mathbb{P}[\mathcal{E}] \ge 1-4n\exp\left(-(n-1)\frac{\delta^2}{3(1-p)}\right). \tag{C.3}$$

Therefore, for any η there exist n_0 such that for any $n \ge n_0$ we have $\mathbb{P}[\mathcal{E}] \ge 1-\eta$. In the rest of the proof we consider a draw of v_{ij}'s that belongs to the event $\mathcal{E}$, which happens with probability at least $1-\eta$. For any node $a \in A$, we let $d_a(A)$ denote the number of nodes $a' \in A$ such that $x_{aa'} = 1$. We also let $d_a(B)$ denote the number of nodes $b \in B$ such that $x_{ab} = 1$.

Claim 1. *For any $\Delta < 1$, there exists a set $S_A \subseteq A$ of size at least $n(1-\Delta)$ such that for all $a \in S_A$, we have $d_a(A) \ge \left(\Delta - (p+\delta)\left(1+\frac{H-L}{H-\gamma}\right)\right)n$ (similar statement holds for B).*

Proof of Claim 1. The claim evidently holds for $\Delta < (p+\delta)(1+\frac{H-L}{H-\gamma})$. We next prove the claim for $\Delta \ge (p+\delta)(1+\frac{H-L}{H-\gamma})$. Suppose the contrary, meaning there exist at least $n\Delta$ individuals in A with degree less than $\left(\Delta-(p+\delta)(1+\frac{H-L}{H-\gamma})\right)n$. The utility of each of these individuals is less than $n\left(\Delta-(p+\delta)(1+\frac{H-L}{H-\gamma})\right)(H-\gamma) + (p+\delta)n(H-\gamma) = n(\Delta(H-\gamma)-(p+\delta)(H-L))$, where the first term is the maximum utility they obtain by connecting to nodes in A and the second term is the maximum utility they obtain by connecting to nodes in B. If all of them deviate altogether and connect to each other, then the utility of each of them is greater than or equal to $(n\Delta - n(p+\delta))(H-\gamma) + n(p+\delta)(L-\gamma)$, where the first term is utility they obtain from high value links and the second term is the utility they obtain from low value links. Since, we have a strong Nash equilibrium, it cannot be the case that all these individuals benefit from this deviation, proving Claim 1.

Claim 2. *For*

$$p < \frac{\min\{\beta\gamma, \gamma-L\}}{2(H-\gamma)+\left(3+\frac{H-L}{H-\gamma}\right)\min\{\beta\gamma,\gamma-L\}}$$

and

$$\Delta = p\left(\frac{H-\gamma}{\min\{\beta\gamma,\gamma-L\}} + 2 + \frac{H-L}{H-\gamma}\right),$$

there exists sufficiently small δ for which none of the individuals in A is connected to S_B and none of the individuals in B is connected to S_A.

Proof of Claim 2. Suppose the contrary, meaning there exists $b \in B$ that is connected to a subset of the agents in S_A. If agent b disconnects from S_A, then his/her marginal utility is lower bounded by

$$-x(H-\gamma) - y(L-\gamma) + \beta\gamma \max\left\{\left(\left(\Delta-(p+\delta)\left(1+\frac{H-L}{H-\gamma}\right)\right)n - x - y\right), 0\right\}, \tag{C.4}$$

where x is the number of high valuation links from b to S_A and y is the number of low valuation links from b to S_A. The minimum of (C.4) is the solution of the following problem.

$$\min -x(H-\gamma) - y(L-\gamma) + \beta\gamma\max\left\{\left(\left(\Delta-(p+\delta)\left(1+\frac{H-L}{H-\gamma}\right)\right)n-x-y\right),0\right\}$$
$$\text{s.t. } x, y \ge 0,\ x+y \le n(1-\Delta),\ x \le (p+\delta)n. \tag{C.5}$$

Given the specified conditions on p and Δ, the solution to (C.5) becomes

$$n\left(-(p+\delta)(H-\gamma)\right) + n\left(\left(\Delta-(p+\delta)\left(1+\frac{H-L}{H-\gamma}\right)-(p+\delta)\right)\min\{\beta\gamma,\gamma-L\}\right),$$

which is positive for small δ (given the specified p and Δ). Thus, there exists sufficiently small δ for which disconnecting from S_A is profitable for b, hence showing Claim 2. Therefore, for any η, there exists n_0 such that for $n \geq n_0$ with probability at least $1-\eta$ there exists subset S_A of A and subset S_B of B each with size at least

$$n\left(1-p\left(\frac{H-\gamma}{\min\{\beta\gamma,\gamma-L\}}+2+\frac{H-L}{H-\gamma}\right)\right)$$

such that no individual from B is connected to S_A and no individual from A is connected to S_B.

This completes the proof of theorem. □

References

Acemoglu, D., Malekian, A., Ozdaglar, A., 2016. Network security and contagion. J. Econ. Theory 166, 536–585.
Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2015. Systemic risk and stability in financial networks. Amer. Econ. Rev. 105 (2), 564–608.
Acquisti, A., Taylor, C., Wagman, L., 2016. The economics of privacy. J. Econ. Lit. 54 (2), 442–492.
Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74 (1), 47–97.
Baerveldt, C., Van Duijn, M.A., Vermeij, L., Van Hemert, D.A., 2004. Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. Soc. Netw. 26 (1), 55–74.
Bala, V., Goyal, S., 1997. Self-Organization in Communication Networks. Econometric Institute Research Papers.
Bala, V., Goyal, S., 2000. A noncooperative model of network formation. Econometrica 68 (5), 1181–1229.
Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's who in networks. Wanted: the key player. Econometrica 74 (5), 1403–1417.
Blalock, H.M., 1982. Race and Ethnic Relations. Prentice Hall, Englewood Cliffs, NJ.
Blau, P.M., 1997. Inequality and Heterogeneity: A Primitive Theory of Social Structure, vol. 7. Free Press, New York.
Blume, L., Easley, D., Kleinberg, J., Kleinberg, R., Tardos, E., 2013. Network formation in the presence of contagious risk. ACM Trans. Econ. Comput. 1 (2), 6–15.
Bramoullé, Y., Kranton, R., 2007. Public goods in networks. J. Econ. Theory 135 (1), 478–494.
Calzolari, G., Pavan, A., 2006. On the optimality of privacy in sequential contracting. J. Econ. Theory 130 (1), 168–204.
Celis, L.E., Mousavifar, A.S., 2015. A model for social network formation: efficiency, stability and dynamics. ArXiv preprint arXiv:1510.09025.
Currarini, S., Jackson, M.O., Pin, P., 2009. An economic model of friendship: homophily, minorities, and segregation. Econometrica 77 (4), 1003–1045.
Davidsen, J., Ebel, H., Bornholdt, S., 2002. Emergence of a small world from local interactions: modeling acquaintance networks. Phys. Rev. Lett. 88 (12), 128701.
Dembo, A., Zeitouni, O., 1998. Large Deviations Techniques and Applications. Springer.
Dutta, B., Mutuswami, S., 1997. Stable networks. J. Econ. Theory 76 (2), 322–344.
Dwork, C., Roth, A., 2013. The algorithmic foundations of differential privacy. Theor. Comput. Sci. 9 (3–4), 211–407.
Easley, D., Kleinberg, J., 2010. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press.
Elliott, M., Golub, B., Jackson, M.O., 2014. Financial networks and contagion. Amer. Econ. Rev. 104 (10), 3115–3153.
Fabrikant, A., Luthra, A., Maneva, E., Papadimitriou, C.H., Shenker, S., 2003. On a network creation game. In: Proceedings of the 22nd Annual Symposium on Principles of Distributed Computing, pp. 347–351.
Fagiolo, G., 2007. Clustering in complex directed networks. Phys. Rev. E 76 (2), 26107.
Fong, E., Isajiw, W.W., 2000. Determinants of friendship choices in multiethnic society. Sociol. Forum 15 (2), 249–271.
Fowler, J.H., Dawes, C.T., Christakis, N.A., 2009. Model of genetic variation in human social networks. Proc. Nat. Acad. Sci. 106 (6), 1720–1724.
Fudenberg, D., Levine, D.K., 1998. The Theory of Learning in Games, vol. 2. MIT Press.
Galeotti, A., Goyal, S., 2010. The law of the few. Amer. Econ. Rev. 100 (4), 1468–1492.
Galeotti, A., Goyal, S., Jackson, M.O., Vega-Redondo, F., Yariv, L., 2010. Network games. Rev. Econ. Stud. 77 (1), 218–244.
Galeotti, A., Goyal, S., Kamphorst, J., 2006. Network formation with heterogeneous players. Games Econ. Behav. 54 (2), 353–372.
Hirshleifer, J., 1980. Privacy: its origin, function, and future. J. Legal Stud. 9 (4), 649–664.
Holme, P., Kim, B.J., 2002. Growing scale-free networks with tunable clustering. Phys. Rev. E 65 (2), 026107.
Hui, K.-L., Png, I.P., 2006. The economics of privacy. Handb. Inform. Syst. Econ. 1 (2), 471–493.
Jackson, M.O., 2003. The stability and efficiency of economic and social networks. Adv. Econ. Des., 319–361.
Jackson, M.O., 2005. A survey of network formation models: stability and efficiency. Group Form. Econ. Netw. Clubs Coalit., 11–49.
Jackson, M.O., Rogers, B.W., 2007. Meeting strangers and friends of friends: how random are social networks? Amer. Econ. Rev. 97 (3), 890–915.
Jackson, M.O., Watts, A., 2002. The evolution of social and economic networks. J. Econ. Theory 106 (2), 265–295.
Jackson, M.O., Wolinsky, A., 1996. A strategic model of social and economic networks. J. Econ. Theory 71 (1), 44–74.
Kearns, M., Roth, A., Zhiwei, S., Yaroslavtsev, G., 2016. Private algorithms for the protected in social network search. Proc. Nat. Acad. Sci. 113 (4), 913–918.
Kempe, D., Kleinberg, J., Tardos, É., 2003. Maximizing the spread of influence through a social network. Theor. Comput. 11 (4), 105–147.
Kleinberg, J., Ligett, K., 2013. Information-sharing in social networks. Games Econ. Behav. 82 (C), 702–716.
Kleinberg, J., Papadimitriou, C.H., Raghavan, P., 2001. On the value of private information. In: Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge, pp. 249–257.
Kossinets, G., Watts, D.J., 2006. Empirical analysis of an evolving social network. Science 311 (5757), 88–90.
Laudon, K.C., 1996. Markets and privacy. Commun. ACM 39 (1), 92–104.
Lazarsfeld, P., Merton, R., Berger, M., Abel, T., Page, C., 1954. Friendship as a social process: a substantive and methodological analysis. Freedom Control Mod. Soc. 18 (1), 18–66.
Leiserson, C.E., Rivest, R.L., Stein, C., Cormen, T.H., 2001. Introduction to Algorithms. McGraw–Hill.
Liang, Y., Poor, H.V., Shamai, S., 2009. Information theoretic security. Found. Trends Commun. Inf. Theor. 5 (4–5), 355–580.
Lori, A., 2012. I Know who You Are and I Saw what You Did: Social Networks and the Death of Privacy. Simon and Schuster.
Magi, T.J., 2011. Fourteen reasons privacy matters: a multidisciplinary review of scholarly literature. Libr. Q. 81 (2), 187–209.
Marsden, P.V., 1987. Core discussion networks of Americans. Amer. Sociol. Rev., 122–131.
Marsden, P.V., 1988. Homogeneity in confiding relations. Soc. Netw. 10 (1), 57–76.
McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27 (1), 415–444.
Medus, A., Dorso, C.O., 2013. Triadic closure mechanism in face-to-face and online social relationship networks. ArXiv preprint, arXiv:1312.3496.
Moody, J., 2001. Race, school integration, and friendship segregation in America. Amer. J. Sociol. 107 (3), 679–716.

Morris, S., 2000. Contagion. Rev. Econ. Stud. 67 (1), 57–78.
Newell, B.C., 2011. Rethinking reasonable expectations of privacy in online social networks. Richmond J. Law Technol. 17 (4), 12–16.
Newman, M.E., 2003. The structure and function of complex networks. SIAM Rev. 45 (2), 167–256.
Newman, M.E., 2004. Coauthorship networks and patterns of scientific collaboration. Proc. Nat. Acad. Sci. 101 (1), 5200–5205.
Patacchini, E., Zenou, Y., 2016. Racial identity and education in social networks. Soc. Netw. 44, 85–94.
Posner, R.A., 1981. The economics of privacy. Amer. Econ. Rev. 71 (2), 405–409.
Rapoport, A., 1953. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. Bull. Math. Biophys. 15 (4), 523–533.
Samuelson, P., 2000. Privacy as intellectual property? Stanf. Law Rev. 52 (5), 1125–1173.
Stigler, G.J., 1980. An introduction to privacy in economics and politics. J. Legal Stud. 9 (4), 623–644.
Tardos, E., Wexler, T., 2007. Network formation games and the potential function method. Algorithmic Game Theory, 487–516.
Taylor, C.R., 2004. Consumer privacy and the market for customer information. RAND J. Econ. 35 (4), 631–650.
Varian, H.R., 2009. Economic aspects of personal privacy. Internet Policy Econ., 101–109.
Vázquez, A., 2003. Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E 67 (5), 056104.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393 (6684), 440–442.
Westin, A.F., 1968. Privacy and freedom. Wash. Lee Law Rev. 25 (1), 166–176.