

Networks in Their Surrounding Contexts

Ana Paula

Fabrício Benevenuto

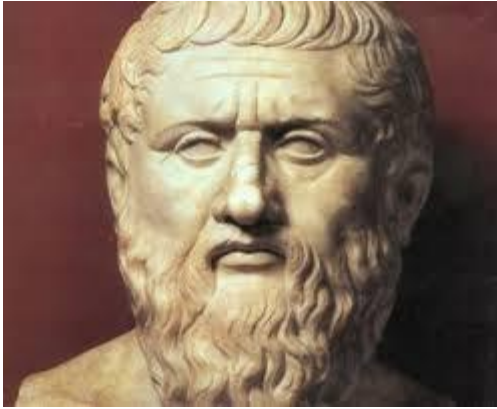
Surrounding Contexts

“... Fatores que existem fora da rede, mas que afetam a estrutura da rede e a forma que ela evolui” (EASLEY and KLEINBERG, 2010, p. 77)

Homofilia

- Capítulo 3: estruturas típicas e processo básico de formação de links
- Formação de links é um processo mais complexo
 - Amizades se forma porque 2 pessoas se conheceram através de um amigo
 - Amizades se forma porque 2 pessoas trabalharam ou estudaram juntas
 - Amizades ainda se formam por vários outros contextos: religião, time de futebol, etc.

Observações antigas relacionadas



Platão

Similaridade leva a amizade



Aristóteles

Pessoas amam aqueles que são como eles

Homophily

“... the principle that we tend to be similar to our friends.” (EASLEY and KLEINBERG, 2010, p. 77)

- Some Dimensions of Similarities
 - Racial and Ethnic
 - Age
 - Places
 - Occupations
 - Interests
 - Beliefs, and so on...

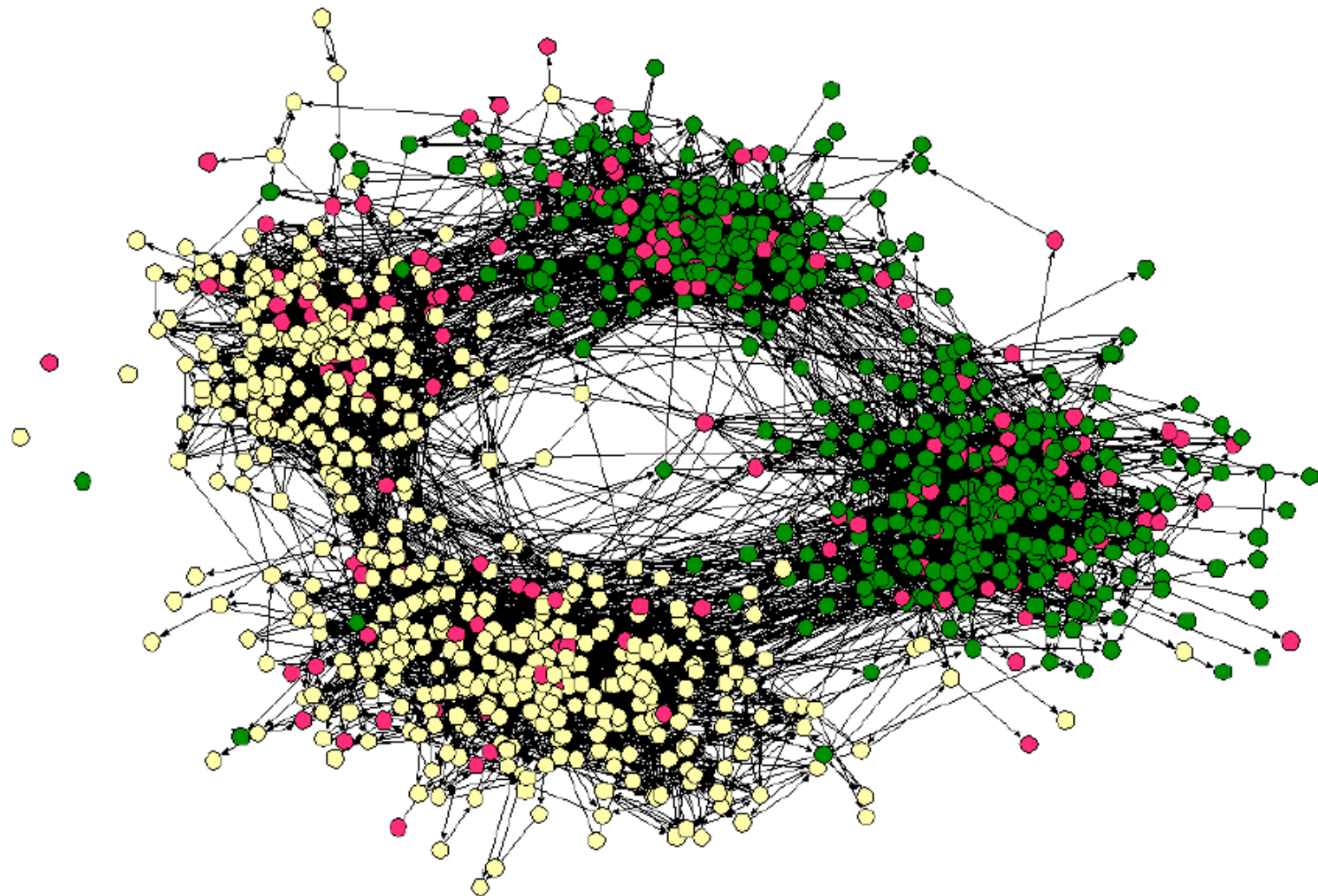


Figure 4.1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town's middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively [304].

Measuring Homophily

Há homofilia nessa rede?

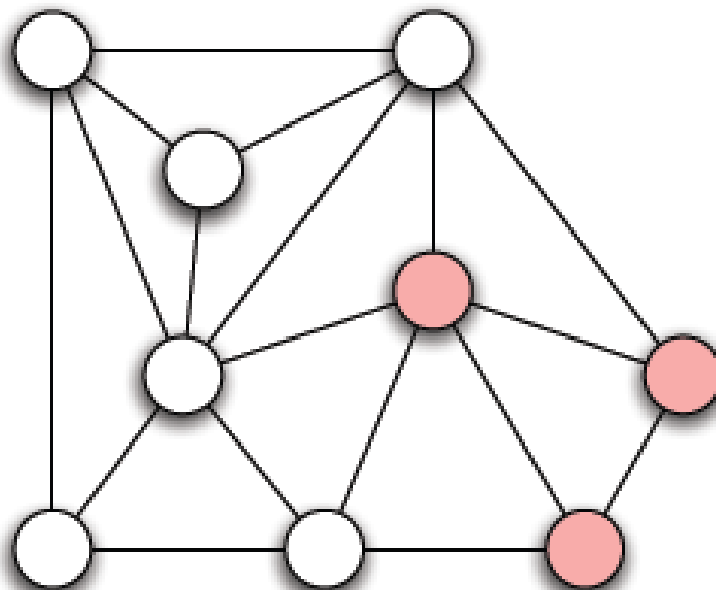


Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

Homofilia

- p = probabilidade de nodo ser homem
- q = probabilidade de nodo ser mulher

Dada uma aresta, a chance dela ser homem-mulher = $pq + qp = 2pq$

Teste de homofilia: Se a fração de arestas entre gêneros for significativamente menor do que $2pq$, há evidência de homofilia na rede

Measuring Homophily

$$p = 3/9 = 2/3$$

$$q = 6/9 = 2/3$$

$$2pq = 4/9 = 8/18$$

Existem 5/18 arestas homen-mulher no grafo

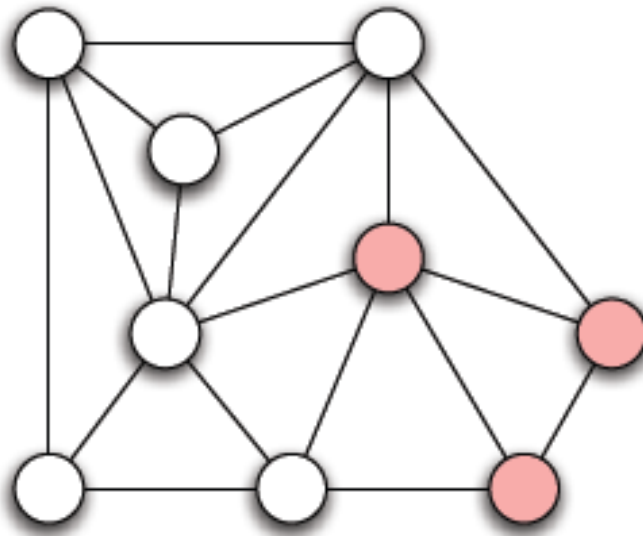


Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

Teste de Homofilia - Considerações

- “Significante menor que $2pq$ ” precisa ser definido
 - Ex. Ordem de grandeza de diferença
 - Se houverem vários grafos pode-se usar intervalo de confiança ou mesmo analisar a distribuição
- Teste de homofilia pode ser generalizado para grafo com mais tipos diferentes de nodos
 - Considera-se aresta homogênea ou heterogênea

Inverse Homophily

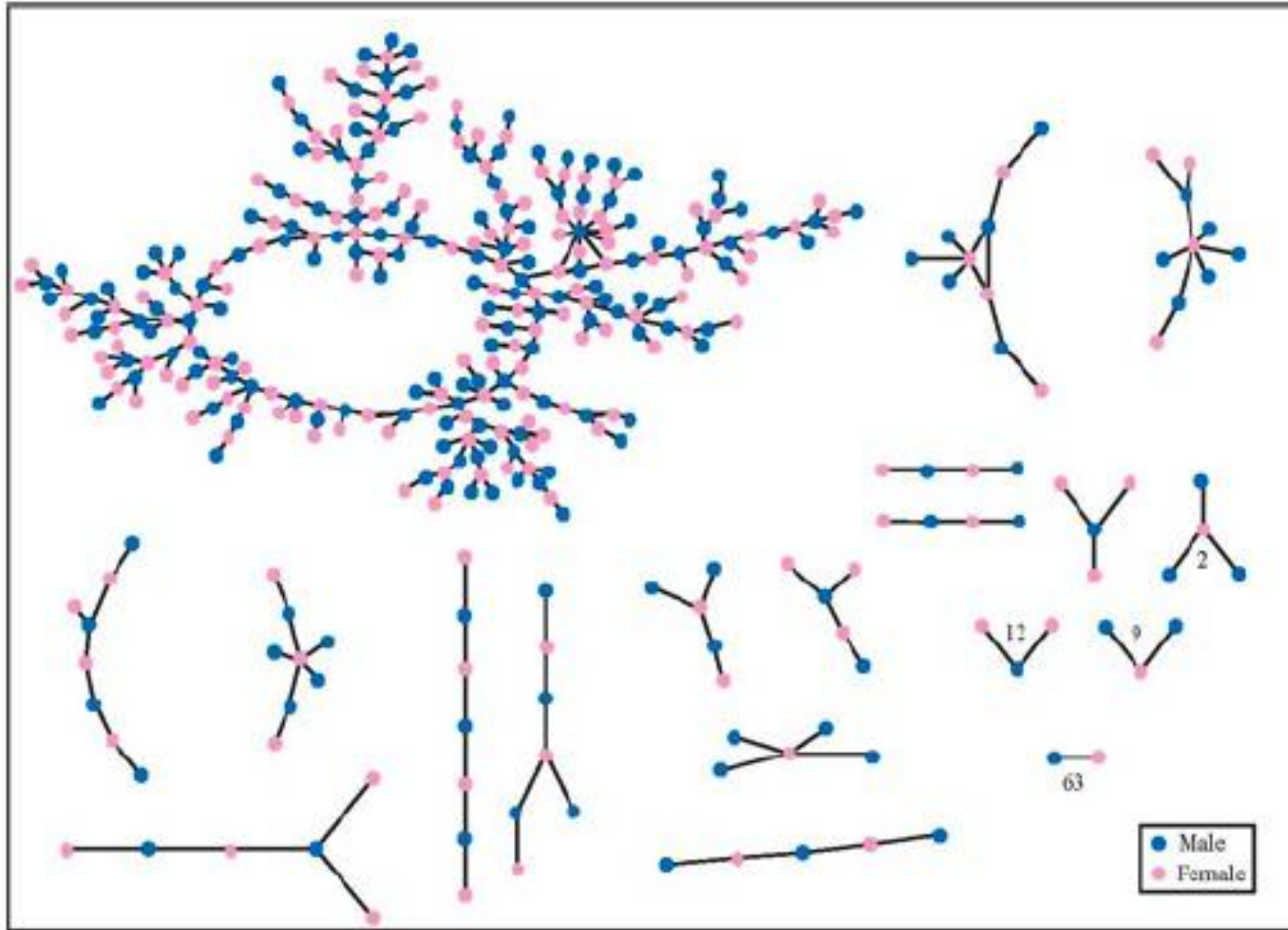


Figure 2.7: A network in which the nodes are students in a large American high school, and an edge joins two who had a romantic relationship at some point during the 18-month period in which the study was conducted [48].

Homofilia e Influência Social

- Homofilia: nodos similares tendem a se conectar
- Influência: Nodos conectados se tornam similares (mudança de comportamento)
- Which is cause and which is effect?
 - Do birds of a feather flock together? (Homophily or associative sorting)
 - Do you change your behavior based on the behavior of your peers? (Social contagion or influence)

Se seu amigo pulou no buraco, você também pularia?

- Sim, ele me inspira.
 - **Influência ou contágio social**
- Sim, na verdade nós nos tornamos amigos por gostarmos de pular em buracos.
 - **Homofilia**
- Sim. Era nossa única opção. Se não pulássemos, nós morreríamos.
 - **Nem homofilia, nem contágio social**

Diferenciando Homofilia e Influência Social

- Homofilia (seleção): pessoas escolhem amigos com algo em comum
- Influência social: amigos influenciam escolhas uns dos outros
- Diferenciar os dois tipos é importante em diversos cenários:
 - Ex. Usuários de droga em uma escola. Convencer um usuário a largar pode impactar o grupo?
 - Ao olhar o snapshot de um grafo, não sabemos o que causou a formação de grupos
- Estudo relacionado a obesidade (análise de 32 anos)
 - Notaram evidências das duas forças

Affiliation Networks

- Grafo bipartido entre usuários e “focal points”
- Permite várias interpretações. Ex. Uma empresa pode influenciar outra?

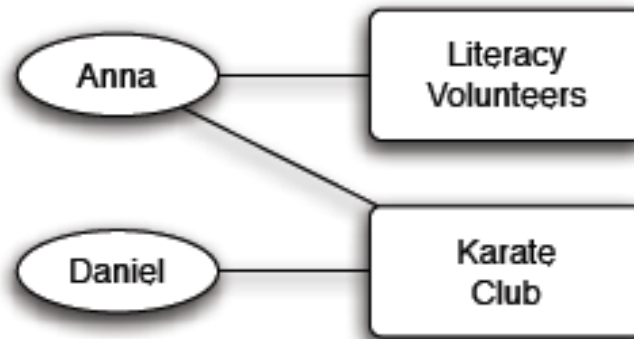


Figure 4.3: An affiliation network is a bipartite graph that shows which individuals are affiliated with which groups or activities. Here, Anna participates in both of the social foci on the right, while Daniel participates in only one.

Affiliation Networks

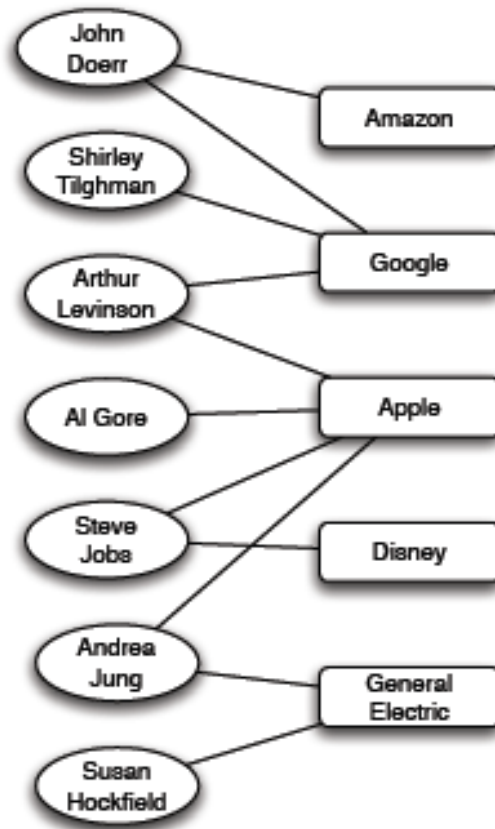


Figure 4.4: One type of affiliation network that has been widely studied is the memberships of people on corporate boards of directors [296]. A very small portion of this network (as of 2009) is shown here. The structural pattern of memberships can reveal subtleties in the interactions among both the board members and the companies.

Social-Affiliation Networks

“Pessoas que participam da mesma afiliação tem chance de se tornarem amigas. Se duas pessoas são amigas, elas podem influenciar umas as outras.”

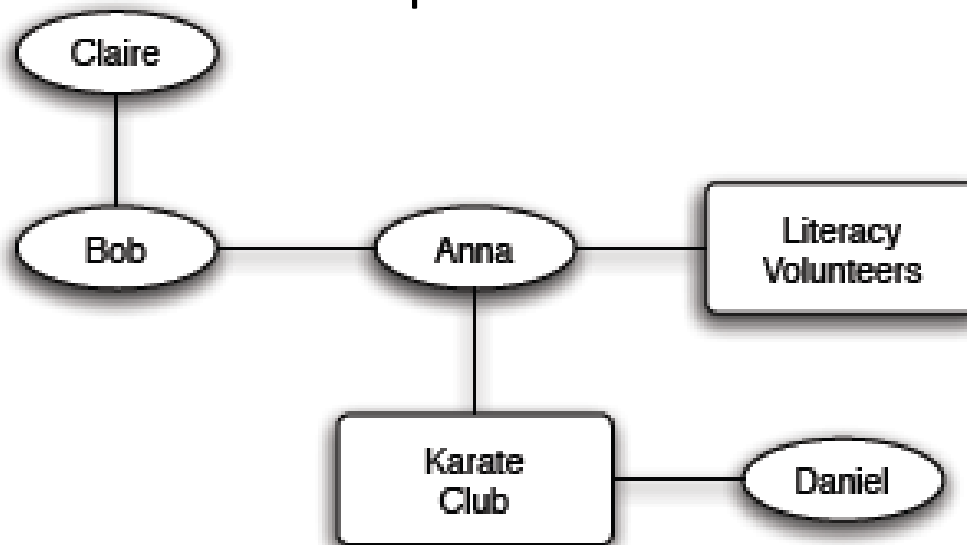
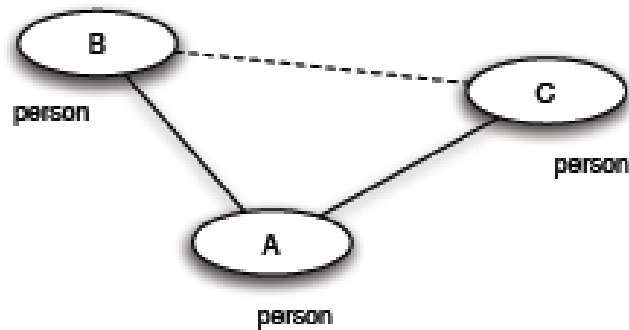
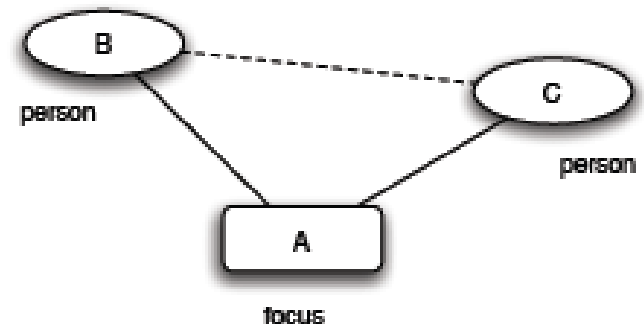


Figure 4.5: A social-affiliation network shows both the friendships between people and their affiliation with different social foci.

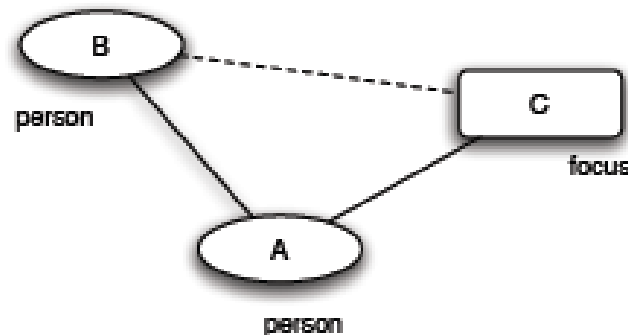
Closure Processes



(a) *Triadic closure*



(b) *Focal closure*



(c) *Social influence*

Figure 4.6: Each of triadic closure, focal closure, and membership influence corresponds to the closing of a triangle in a social-affiliation network.

Closure Processes

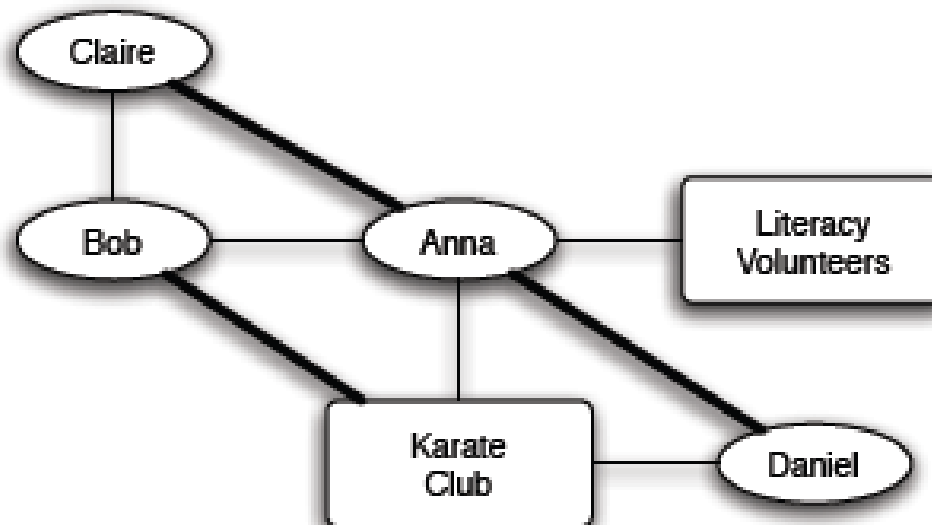


Figure 4.7: In a social-affiliation network containing both people and foci, edges can form under the effect of several different kinds of closure processes: two people with a friend in common, two people with a focus in common, or a person joining a focus that a friend is already involved in.

Formação de links em dados online

Triadic closure

- O quanto maior é a chance de formação de link entre duas pessoas em uma rede social se elas possuem um amigo em comum?
- O quanto maior é essa chance se houverem múltiplos amigos em comum?

Triadic Closure

- (i) Pegue dois snapshots da rede em tempos diferentes
- (ii) Para cada k , identifique todos os pares de nodos possuem exatamente k amigos em comum no primeiro snapshot e que não estão diretamente conectados por uma aresta
- (iii) Defina $T(k)$ como a fração desses pares que formaram um par no segundo snapshot. Essa é estimativa empírica para a probabilidade de que um link se formará entre duas pessoas com k amigos em comum
- (iv) Plote $T(k)$ em função de k para ilustrar o efeito do número de amigos na formação de links

Triadic Closure

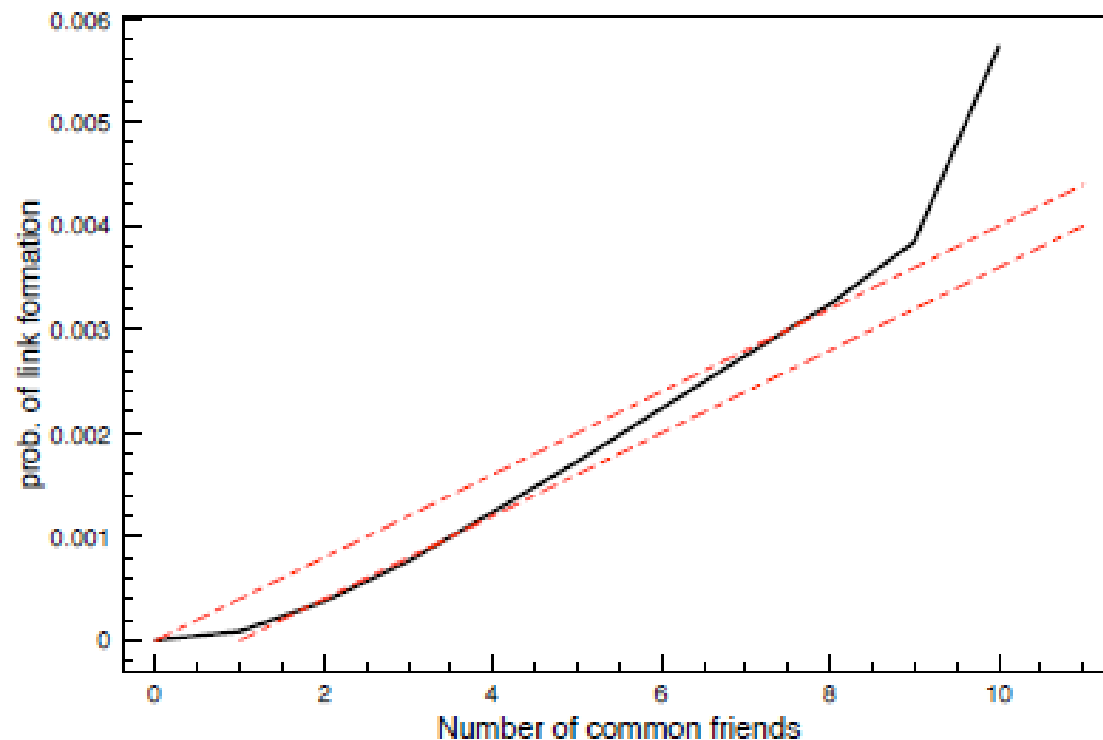


Figure 4.9: Quantifying the effects of triadic closure in an e-mail dataset [255]. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation.

Triadic Closure

- Cada amigo em comum que duas pessoas possuem dão uma probabilidade p de formação de link de um dia pro outro
- $1-p$ = probabilidade de falha – não formação do link
- Se existem k amigos em comum, a probabilidade de falha é $(1-p)^k$, assumindo que k é independente
- $T(k) = 1 - (1-p)^k$ (probabilidade de formação do link)

Triadic Closure

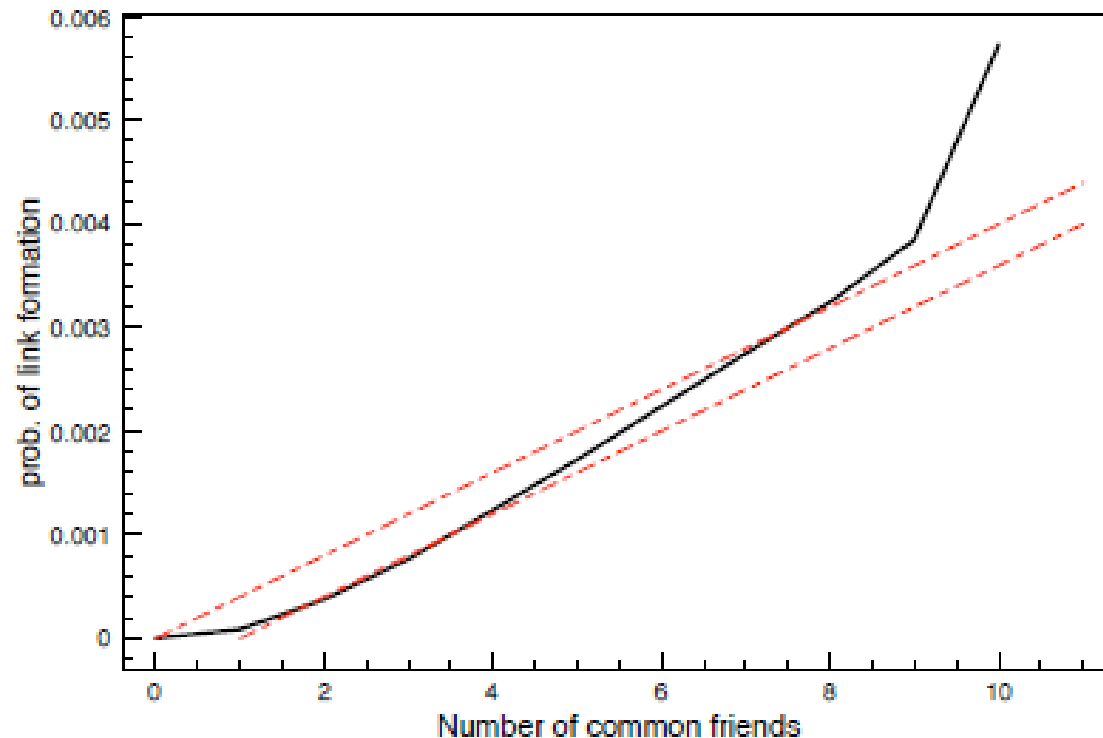


Figure 4.9: Quantifying the effects of triadic closure in an e-mail dataset [255]. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation.

Formação de links em dados online

Focal closure

- **Qual a probabilidade de duas pessoas formarem um link em função do número de foci em comum em que elas estão afiliadas**

Foci = disciplinas realizadas juntos

Focal Closure

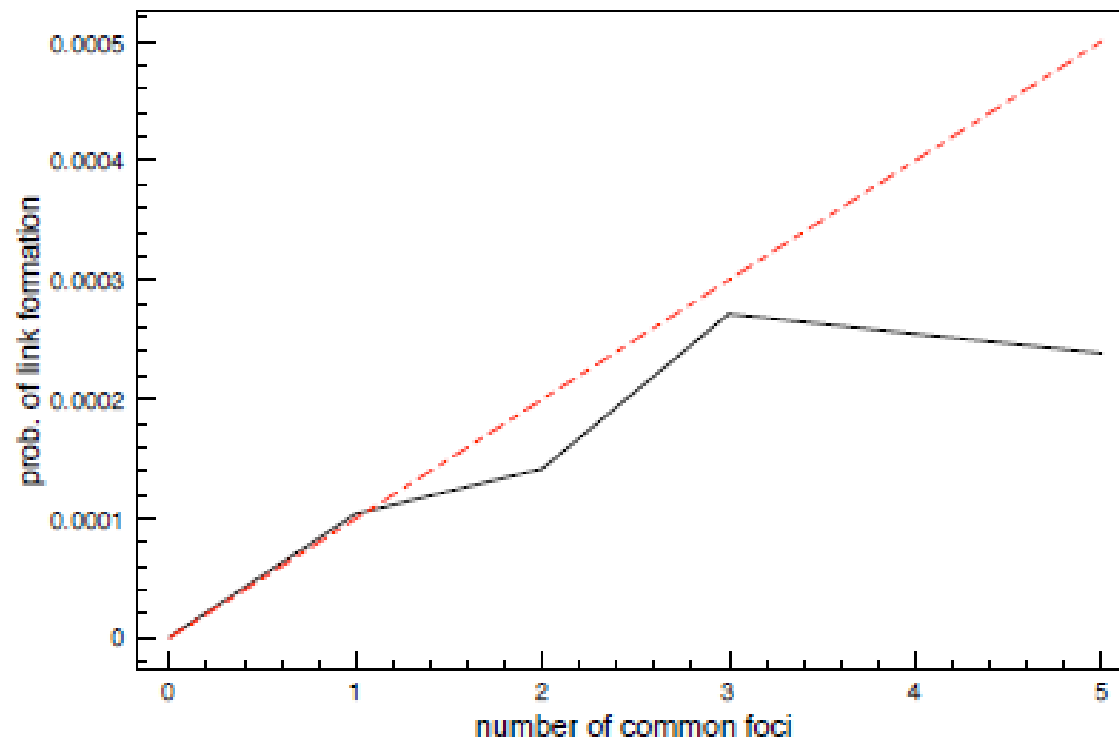


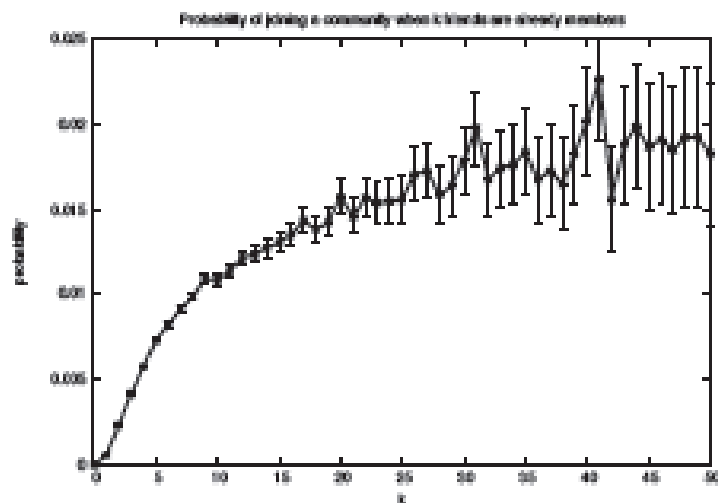
Figure 4.10: Quantifying the effects of focal closure in an e-mail dataset [255]. Again, the curve determined from the data is shown in the solid black line, while the dotted curve provides a comparison to a simple baseline.

Formação de links em dados online

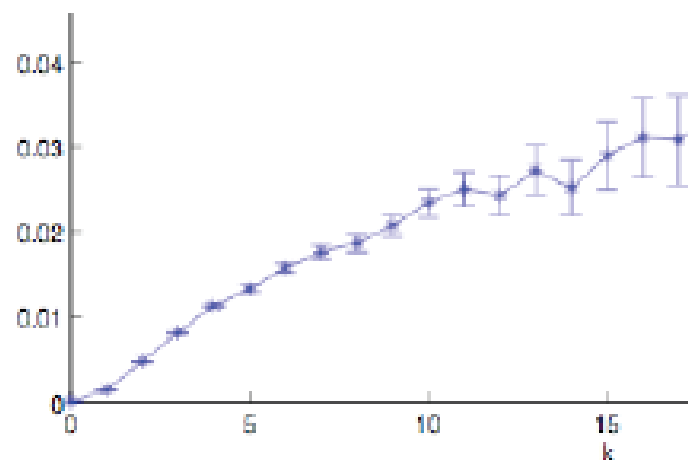
Membership closure

- **Qual a probabilidade de uma pessoa se envolver em uma afiliação em função do número de amigos que já estão envolvidos?**

Membership Closure



(a) *Probability of joining a LiveJournal community as a function of the number of friends who are already members.*



(b) *Probability of editing a Wikipedia articles as a function of the number of friends who have already done so.*

Figure 4.11: Quantifying the effects of membership closure in two large on-line datasets [32, 121].

Quantifying the Interplay Between Selection (homophily) and Social Influence

Considering Wikipedia data discussed before, *does the homophily arise because editors are forming connections with those who have edited the same articles (selection), or is it because editors are led to the articles of those they talk to (social influence)?*

Quantifying the Interplay Between Selection and Social Influence

- Similaridade medida em termos do número de páginas editadas por 2 usuários – interseção/união
- Tempo normalizado em função da primeira vez que houve comunicação
- Similaridade média de todos os pares
- Baseline são os editores que não se comunicaram

Quantifying the Interplay Between Selection and Social Influence

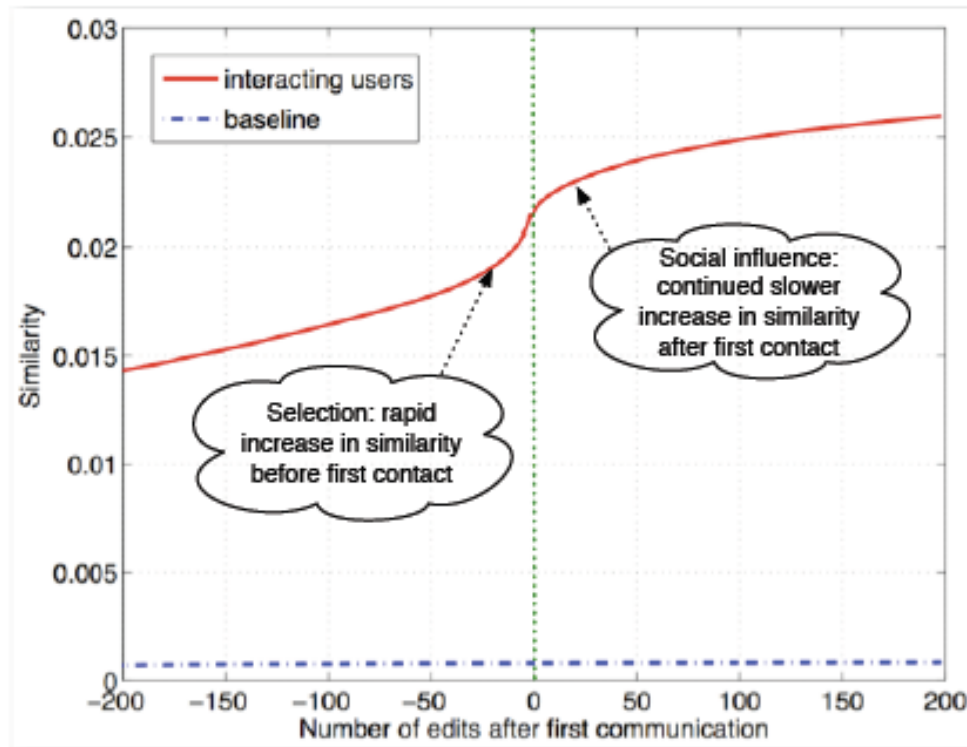


Figure 4.12: The average similarity of two editors on Wikipedia, relative to the time (0) at which they first communicated [121]. Time, on the x -axis, is measured in discrete units, where each unit corresponds to a single Wikipedia action taken by either of the two editors. The curve increases both before and after the first contact at time 0, indicating that both selection and social influence play a role; the increase in similarity is steepest just before time 0.

A Spatial Model of Segregation

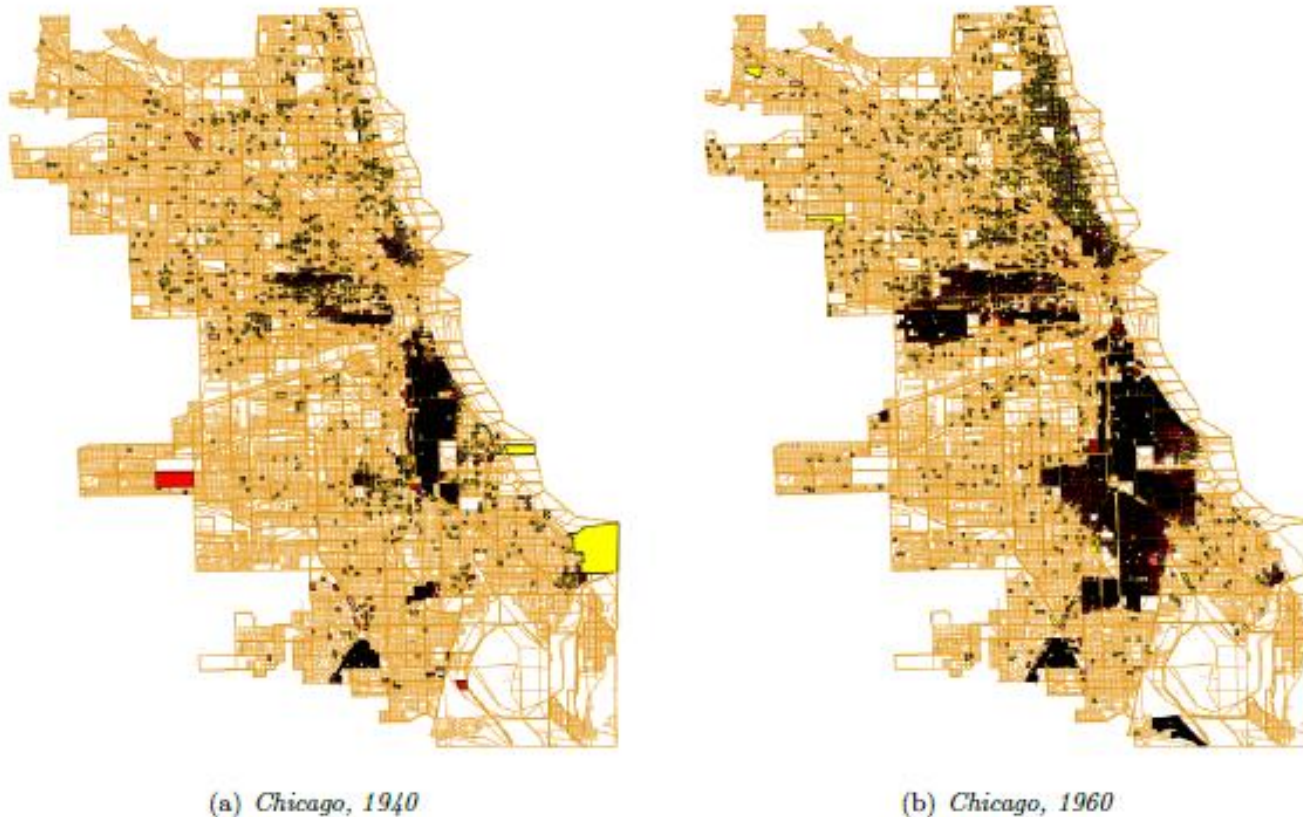


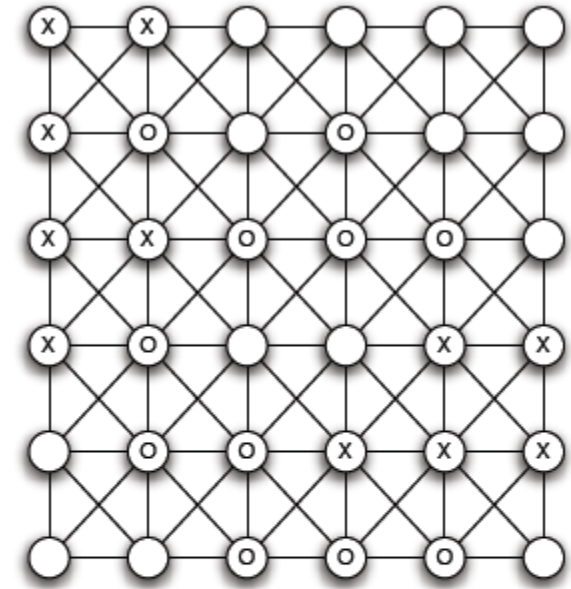
Figure 4.13: The tendency of people to live in racially homogeneous neighborhoods produces spatial patterns of segregation that are apparent both in everyday life and when superimposed on a map — as here, in these maps of Chicago from 1940 and 1960 [297]. In blocks colored yellow and orange the percentage of African-Americans is below 25, while in blocks colored brown and black the percentage is above 75.

A Spatial Model of Segregation

Threshold $t = 3$ – menos de 3 vizinhos do mesmo tipo, o nodo está insatisfeito

X	X				
X	O		O		
X	X	O	O	O	
X	O			X	X
	O	O	X	X	X
		O	O	O	

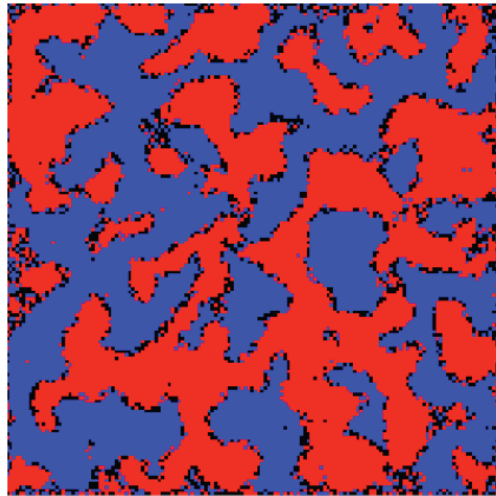
(a) *Agents occupying cells on a grid.*



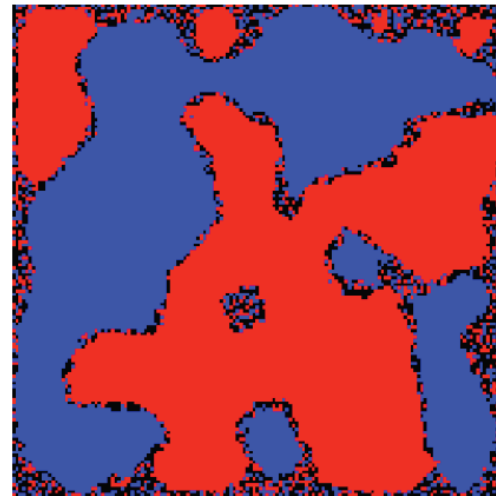
(b) *Neighbor relations as a graph.*

Figure 4.15: In Schelling's segregation model, agents of two different types (X and O) occupy cells on a grid. The neighbor relationships among the cells can be represented very simply as a graph. Agents care about whether they have at least some neighbors of the same type.

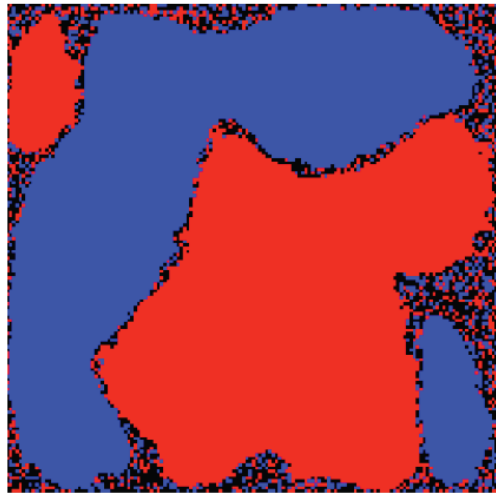
A Spatial Model of Segregation



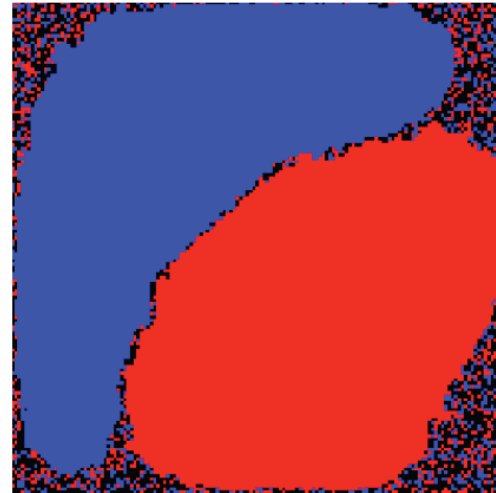
(a) *After 20 steps*



(b) *After 150 steps*



(c) *After 350 steps*



(d) *After 800 steps*

Figure 4.19: Four intermediate points in a simulation of the Schelling model with a threshold t of 4, on a 150-by-150 grid with 10,000 agents of each type. As the rounds of movement progress, large homogeneous regions on the grid grow at the expense of smaller, narrower regions.

É possível inferir o comportamento de pessoas?

- **You are who you know: Inferring user profiles in Online Social Networks**
Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. *ACM International Conference of Web Search and Data Mining (WSDM'10)*, New York, NY, February 2010.
- Thomas, Kurt, Chris Grier, and David M. Nicol. "unfriendly: Multi-party privacy risks in social networks." *Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 2010.