Web-graphs. Lecture 5.

Clustering coefficient in the Buckley-Ostgus model

Let $m \geq 1$, $H^n_{a,m}$ is the random graph the Buckley-Ostgus model

- **Theorem (Eggemann, Noble)**. If $a > 1$ then

$$\mathbb{E}(\#(K_3, H^n_{a,m})) \asymp \ln n, \quad n \to \infty$$

- **Theorem (Eggemann, Noble)**. If $a > 1$ then

$$\mathbb{E}(\#(P_2, H^n_{a,m})) \asymp n, \quad n \to \infty$$

- **Theorem (Eggemann, Noble)**. If $a > 1$ then

$$\mathbb{E}(T(H^n_{a,m})) \asymp \frac{\ln n}{n}, \quad n \to \infty$$

Clustering coefficient in the Buckley-Ostgus model

Let $m \geq 1$, $H^n_{a,m}$ be the random graph the Buckley-Ostgus model

- **Theorem (Tilga)**. Let $m \geq 2,\ a < 1,\ \lambda = \frac{1}{a+1}$. Let P_l be a path of length l. Then while $n\ to \infty$

$$\mathbb{E}(\#(P_l, H^n_{a,m})) = \begin{cases} n^{(2\lambda-1)k+1}\Theta(m^l), & l = 2k \\ n^{(2\lambda-1)k+1}\ln n\Theta(m^l), & l = 2k+1 \end{cases}$$

- **Theorem (Tilga)**. Let $m \geq 2,\ a < 1,\ \lambda = \frac{1}{a+1}$. Let C_l be a cycle of length l. Then while $n\ to \infty$

$$\mathbb{E}(\#(C_l, H^n_{a,m})) = \begin{cases} n^{(2\lambda-1)k}\Theta(m^l), & l = 2k \\ n^{(2\lambda-1)k}\ln n\Theta(m^l), & l = 2k+1 \end{cases}$$

Clustering coefficient in the Buckley-Ostgus model

Let $m \geq 1$, $H^n_{a,m}$ be the random graph the Buckley-Ostgus model

- **Theorem (Tilga)**. Let $m \geq 2,\ a < 1,\ \lambda = \frac{1}{a+1}$. Let K_k be a clique on k nodes, $4 \leq k \leq m+1$ Then while $n\ to \infty$

$$\mathbb{E}(\#(K_k, H^n_{a,m})) = \begin{cases} n^{1+(\lambda-1)(k-1)}\Theta(m^{C_k^2}), & a < \frac{1}{k-2} \\ \ln n\Theta(m^{C_k^2}), & a = \frac{1}{k-2} \\ \Theta(m^{C_k^2}), & a > \frac{1}{k-2} \end{cases}$$

- $a = 1/3$ (≈ 0.27), $\lambda = 3/4$:
  - K_5: $\ln n$
  - K_4: $\sqrt[4]{n}$
  - K_6: const
- $\longrightarrow$ Buckley-Ostgus model is much more realistic than the one of Bollobas-Riordan.

Clustering coefficient in the Buckley-Ostgus model

Let $m \geq 1$, $H^n_{a,m}$ is the random graph the Buckley-Ostgus model

- **Theorem (Tilga)**. Let $m \geq 2,\ a < 1,\ \lambda = \frac{1}{a+1}$. Let $K_{k,l}$ be a complete bipartite graph on k and l nodes in the respective parts. $2 \leq l \leq \min(k, m)$ Then while $n\ to \infty$

$$\mathbb{E}(\#(K_{k,l}, H^n_{a,m})) = \begin{cases} n^{k(1+(\lambda-1)l)}\Theta(m^{kl}), & a < \frac{1}{l-1} \\ (\ln n)^k \Theta(m^{kl}), & a = \frac{1}{l-1} \\ \Theta(m^{kl}), & a > \frac{1}{l-1} \end{cases}$$

- $a = 1/3$ (≈ 0.27), $\lambda = 3/4$:
  - $K_{k,4}$ $(l = 4)$: $(\ln n)^k$
  - $K_{k,3}$ $(l = 3)$: $n^{k/4}$
  - $K_{k,5}$: almost not exists

The Strength of Weak Ties

Granovetter's analysis

- Main reason for the complex network analysis – possibility to understand how global complex events and properties arise from the level of individual nodes and links.
- How information flows through a social network, how different nodes can play structurally distinct roles in this process, and how these structural considerations shape the evolution of the network itself over time?
- As part of his Ph.D. thesis research in the late 1960s, Mark Granovetter interviewed people who had recently changed employers to learn how they discovered their new jobs
- In keeping with earlier research, he found that many people learned information leading to their current jobs through personal contacts.
- But perhaps more strikingly, these personal contacts were often described by interview subjects as acquaintances rather than close friends.
- This is a bit surprising: your close friends presumably have the most motivation to help you when you're between jobs, so why is it so often your more distant acquaintances who are actually to thank for crucial information leading to your new job?

The Strength of Weak Ties

Bridges and Local Bridges

- Information about good jobs is something that is relatively scarce; hearing about a promising job opportunity from someone suggests that they have access to a source of useful information that you don't.
- We analyze node A. It has four friends: C, D, E and B.
- Friend B significantly different from others

From Easly & Kleinberg, Networks, Crowds and Markets

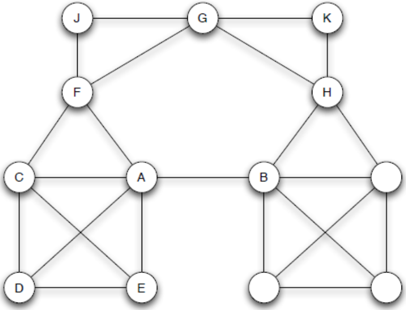


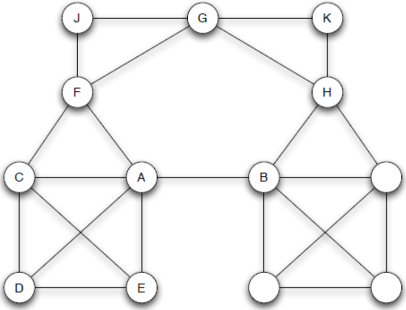
Figure 3.4: The A-B edge is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4.

The Strength of Weak Ties

Bridges and Local Bridges

- A's links to C, D, and E connect her to a tightly-knit group of friends who all know each other
- The link to B seems to reach into a different part of the network
- We could speculate, then, that the structural peculiarity of the link to B will translate into differences in the role it plays in A's everyday life: while the tightly-knit group of nodes A, C, D, and E will all tend to be exposed to similar opinions and similar sources of information, A's link to B offers her access to things she otherwise wouldn't necessarily hear about.

From Easly & Kleinberg, Networks, Crowds and Markets

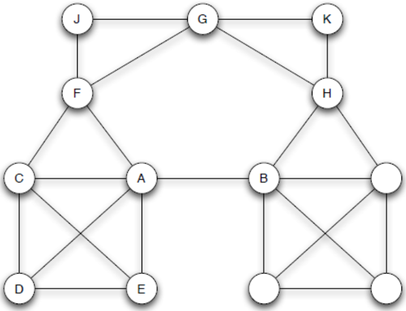


The Strength of Weak Ties

Bridges and Local Bridges

- We say that an edge joining two nodes A and B in a graph is a bridge if deleting the edge would cause A and B to lie in two different components. In other words, this edge is literally the only route between its endpoints, the nodes A and B.
- Bridges are very rare in real networks
- In the example the A-B edge isn't the only path that connects its two endpoints; though they may not realize it, A and B are also connected by a longer path through F, G, and H.

From Easly & Kleinberg, Networks, Crowds and Markets



The Strength of Weak Ties

Bridges and Local Bridges

- We say that an edge joining two nodes A and B in a graph is a local bridge if its endpoints A and B have no friends in common — in other words, if deleting the edge would increase the distance between A and B to a value strictly more than two. $\longrightarrow$ Absence of triadic closure.
- We say that the span of a local bridge is the distance its endpoints would be from each other if the edge were deleted

From Easly & Kleinberg, Networks, Crowds and Markets

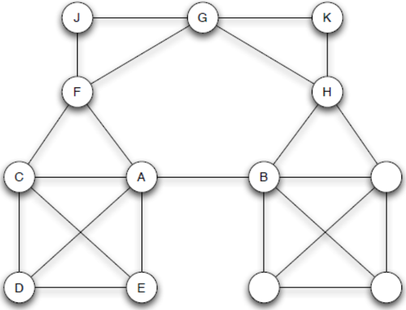


Figure 3.4: The A-B edge is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4.

- Local bridges, especially those with reasonably large span, still play roughly the same role that bridges do, though in a less extreme way — they provide their endpoints with access to parts of the network, and hence sources of information, that they would otherwise be far away from.
- This is a first network context in which to interpret Granovetter's observation about job-seeking: we might expect that if a node like A is going to get truly new information, the kind that leads to a new job, it might come unusually often (though certainly not always) from a friend connected by a local bridge.

The Strength of Weak Ties

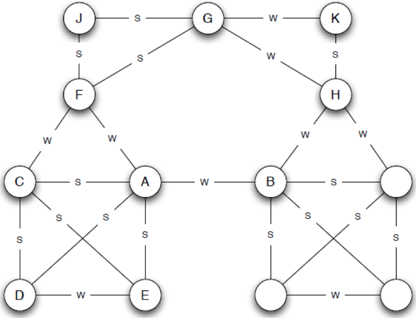
- Of course, Granovetter's interview subjects didn't say, "I learned about the job from a friend connected by a local bridge."
- We should introduce the notion of link's *strength*.
- We'll categorize all links in the social network as belonging to one of two types: **strong ties** (the stronger links, corresponding to friends), and **weak ties** (the weaker links, corresponding to acquaintances).

The Strength of Weak Ties

The Strong Triadic Closure Property

- An example of annotated network
- If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if A's edges to B and C are both strong ties.
- Clearly the Strong Triadic Closure Property is too extreme for us to expect it hold across all nodes of a large social network.
- The Strong Triadic Closure Property $\longrightarrow$ local bridges are most likely weak ties.

From Easley & Kleinberg, Networks, Crowds and Markets



Real world examples

Who-talks-to-whom network

- "Who-talks-to-whom" data exhibits the two ingredients we need for empirical evaluation of hypotheses about weak ties: it contains the network structure of communication among pairs of people, and we can use the total time that two people spend talking to each other as a proxy for the strength of the tie.
- Onnela et al. Structure and tie strengths in mobile communication networks. Proc. Natl. Acad. Sci. USA, 104:7332–7336, 2007
- The nodes correspond to cell-phone users, and there is an edge joining two nodes if they made phone calls to each other in both directions over an 18-week observation period.
- Definition of the "smooth" local bridges: neighborhood overlap

$$\frac{\text{number of nodes who are neighbors of } \textit{both} \text{ A and B}}{\text{number of nodes who are neighbors of } \textit{at least one of} \text{ A and B}}$$

- Local bridges are the edges of neighborhood overlap 0 — and hence we can think of edges with very small neighborhood overlap as being "almost" local bridges

Real world examples

Who-talks-to-whom network

From Easley & Kleinberg, Networks, Crowds and Markets

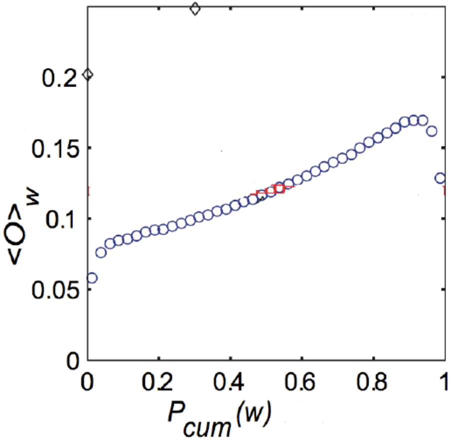


Figure 3.7: A plot of the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength. The fact that overlap increases with increasing tie strength is consistent with the theoretical predictions from Section 3.2. (Image from [334].)

Real world examples

Who-talks-to-whom network

- Onnela et al. proposed one more way to check the hypothesis that weak ties serve to link together different tightly-knit communities that each contain a large number of stronger ties.
- They first deleted edges from the network one at a time, starting with the strongest ties and working downward in order of tie strength. $\longrightarrow$ The giant component shrank steadily as they did this, its size going down gradually due to the elimination of connections among the nodes.
- They then tried the same thing, but starting from the weakest ties and working upward in order of tie strength. $\longrightarrow$ In this case, they found that the giant component shrank more rapidly, and moreover that its remnants broke apart abruptly once a critical number of weak ties had been removed.
- This is consistent with a picture in which the weak ties provide the more crucial connective structure for holding together disparate communities, and for keeping the global structure of the giant component intact.

Real world examples

The strength of the Facebook

From Easley & Kleinberg, Networks, crowds and markets

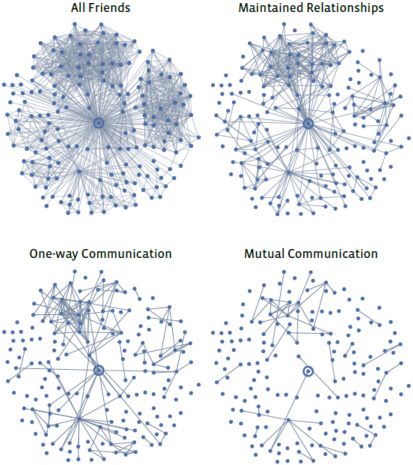


Figure 3.8: Four different views of a Facebook user's network neighborhood, showing the structure of links coresponding respectively to all declared friendships, maintained relationships, one-way communication, and reciprocal (i.e. mutual) communication. (Image from [286].)

Real world examples
The strength of the Twitter

From Easley & Kleinberg, Networks, crowds and markets

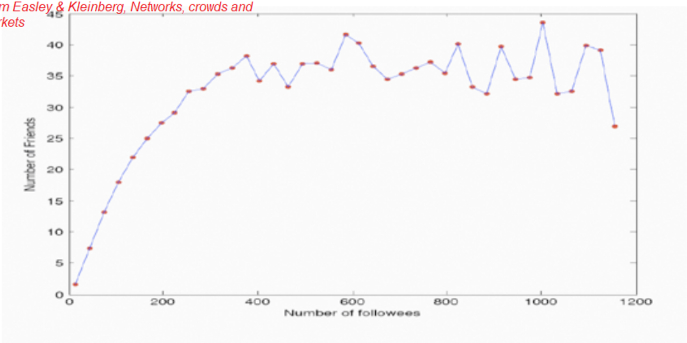


Figure 3.10: The total number of a user's strong ties (defined by multiple directed messages) as a function of the number of followees he or she has on Twitter. (Image from [222].)

Real world examples

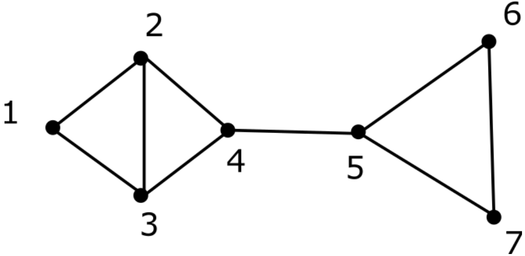
- By definition, each strong tie requires the continuous investment of time and effort to maintain, and so even people who devote a lot of their energy to building strong ties will eventually reach a limit — imposed simply by the hours available in a day — on the number of ties that they can maintain in this way.
- The formation of weak ties is governed by much milder constraints — they need to be established at their outset but not necessarily maintained continuously — and so it is easier for someone to accumulate them in large numbers.
- The strength of such environments like Facebook and Twitter is in simplicity of the weak ties formation.

Betweenness centrality

- Betweenness centrality of a node is a fraction of other nodes shortest paths going through this fixed node
- $P(u, v)$ is the set of shortest paths between nodes u and v ($u \neq v$)
- $P(u, v, t)$ is the set of shortest paths between nodes u and v ($u \neq v$) which go through the node t ($t \neq u$, $t \neq v$)
- Node t betweenness centrality $b(t)$ is defined as follows:

$$b(t) = \sum_{u \neq t, v \neq t} \frac{|P(u, v, t)|}{|P(u, v)|}$$

- Calculate $b(4)$ and $b(2)$ in the following graph



- $b(4) = 9,\ b(2) = 2$