Great Ideas in Computing

University of Toronto CSC196
Fall 2025

Class 23: November 24 (2025)

Announcements and Agenda

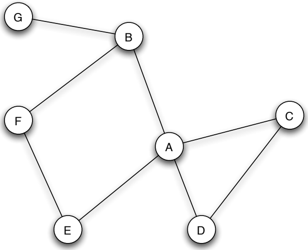
Announcements

- There are no slides for the guest presentation on Wednesday, November 19 by Jonathan Panuelos.
- Assignment 4 is complete and due this Wednesday, November 26 at 3PM
- The second and final quiz will be held Friday, November 28 in the usual tutorial room BA 2139.

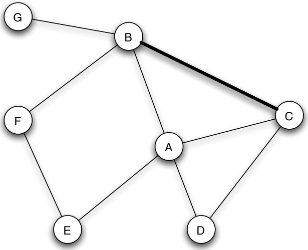
Today's Agenda

- Continue Social networks: Triadic closure, 6 degrees of separation

Triadic closure (undirected graphs)



(a) *Before B-C edge forms.*



(b) *After B-C edge forms.*

Figure: The formation of the edge between B and C illustrates the effects of triadic closure, since they have a common neighbor A. [E&K Figure 3.1]

- Triadic closure: mutual "friends" of say A are more likely (than "normally") to become friends over time.
- How do we measure the extent to which triadic closure is occurring?
- How can we know why a new friendship tie is formed? (Friendship ties can range from "just knowing someone" to "a true friendship" .)

Measuring the extent of triadic closure

- The clustering coefficient of a node A is a way to measure (over time) the extent of triadic closure (perhaps without understanding why it is occurring).

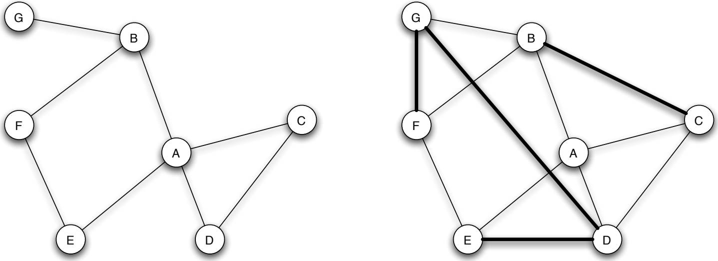
- Let E be the set of an undirected edges of a network graph. (Forgive the abuse of notation where in the previous and next slide E is a node name.) For a node A, the clustering coefficient is the following ratio:

$$\frac{\left|\left\{(B,C) \in E : (B,A) \in E \text{ and } (C,A) \in E\right\}\right|}{\left|\left\{\{B,C\} : (B,A) \in E \text{ and } (C,A) \in E\right\}\right|}$$

- The numerator is the number of all edges (B,C) in the network such that B and C are adjacent to (i.e. mutual friends of) A.

- The denominator is the total number of all unordered pairs $\{B,C\}$ such that B and C are adjacent to A.

Example of clustering coefficient



(a) *Before new edges form.* (b) *After new edges form.*

- The clustering coefficient of node A in Fig. (a) is $1/6$ (since there is only the single edge (C, D) among the six pairs of friends: $\{B, C\}$, $\{B, D\}$, $\{B, E\}$, $\{C, D\}$, $\{C, E\}$, and $\{D, E\}$).
- The clustering coefficient of node A in Fig. (b) increased to $1/2$ (because there are three edges (B, C), (C, D), and (D, E)).
  Note: The new edge $\{D, G\}$ was not created due to triadic closure. Without further information we cannot tell why an edge (e.g., a friendship) was created.

Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

- Bearman and Moody [2004] reported finding that a low clustering coefficient amongst teenage girls implies a higher probability of contemplating suicide (compared to those with high clustering coeficient). Note:The value of the clustering coefficient is also referred to as the *intransitivity coefficient*.

- They report that " Social network effects for girls overwhelmed other variables in the model and appeared to play an unusually significant role in adolescent female suicidality. These variables did not have a significant impact on the odds of suicidal ideation among boys. "

How can we understand these findings?

Bearman and Moody study continued

- Triadic closure (or lack thereof) can provide some plausible explanation.

Bearman and Moody study continued

- Triadic closure (or lack thereof) can provide some plausible explanation.
  Increased opportunity, trust, incentive ; it can be awkward to have friends (especially good friends with strong ties) who are not themselves friends.

Bearman and Moody study continued

- Triadic closure (or lack thereof) can provide some plausible explanation.
  Increased opportunity, trust, incentive ; it can be awkward to have friends (especially good friends with strong ties) who are not themselves friends.
  As far as I can tell, no conclusions are being made about why there is such a difference in gender results.

The study by Bearman and Moody is quite careful in terms of identifying many possible factors relating to suicidal thoughts. Clearly there are many factors involved but the fact that network structure is identified as such an important factor is striking.

Bearman and Moody factors relating to suicidal thoughts

TABLE 3—Logistic Regression of Suicide Attempts, Among Adolescents With Suicidal Ideation, on Individual, School, Family and Network Characteristics

	Suicide Attempts, OR (95% CI)	
	Males	Females
Demographic		
Age	0.956 (0.808, 1.131)	0.920 (0.810, 1.046)
Race/ethnicity		
Black	0.872 (0.414, 1.839)	1.086 (0.680, 1.736)
Other	1.069 (0.662, 1.728)	1.134 (0.810, 1.586)
Socioeconomic status	0.948 (0.872, 1.031)	1.008 (0.951, 1.069)
School and community		
Junior high school	1.588 (0.793, 3.180)	1.271 (0.811, 1.993)
Relative density	0.049 (0.005, 0.521)	0.415 (0.086, 1.996)
Plays team sport	0.985 (0.633, 1.532)	1.020 (0.763, 1.364)
Attachment to school	1.079 (0.823, 1.414)	1.066 (0.920, 1.235)
Religion		
Church attendance	0.975 (0.635, 1.496)	0.818 (0.618, 1.082)
Family and household		
Parental distance	0.925 (0.681, 1.256)	0.955 (0.801, 1.139)
Social closure	1.004 (0.775, 1.299)	0.933 (0.781, 1.115)
Stepfamily	1.058 (0.617, 1.814)	1.368 (0.967, 1.935)
Single-parent household	1.142 (0.698, 1.866)	1.117 (0.800, 1.560)
Gun in household	1.599 (1.042, 2.455)	1.094 (0.800, 1.494)
Family member attempted suicide	1.712 (0.930, 3.150)	1.067 (0.689, 1.651)
Network		
Isolation	0.767 (0.159, 3.707)	1.187 (0.380, 3.708)
Intransitivity index	0.444 (0.095, 2.085)	1.076 (0.373, 3.103)
Friend attempted suicide	1.710 (1.095, 2.671)	1.663 (1.253, 2.207)
Trouble with people	1.107 (0.902, 1.357)	1.119 (0.976, 1.284)
Personal characteristics		
Depression	1.160 (0.960, 1.402)	1.130 (0.997, 1.281)
Self-esteem	1.056 (0.777, 1.434)	0.798 (0.677, 0.942)
Drunkenness frequency	1.124 (0.962, 1.312)	1.235 (1.115, 1.368)
Grade point average	0.913 (0.715, 1.166)	0.926 (0.781, 1.097)
Sexually experienced	1.323 (0.796, 2.198)	1.393 (0.990, 1.961)
Homosexual attraction	1.709 (0.921, 3.169)	1.248 (0.796, 1.956)
Forced sexual relations		1.081 (0.725, 1.613)
No. of fights	0.966 (0.770, 1.213)	1.135 (0.983, 1.310)
Body mass index	0.981 (0.933, 1.032)	1.014 (0.982, 1.047)
Response profile (n = 1/n = 0)	139/493	353/761
F statistic	1.84 (P = .0170)	2.88 (P < .0001)

Note. OR = odds ratio; CI = confidence interval. Logistic regressions; standard errors corrected for sample clustering and stratification on the basis of region, ethnic mix, and school type and size.

The Small World Phenomena

I already mentioned the small worlds phenomena. A mathematical explanation of this phenomiena (expecially how one hones in on a target recipient) was given by J. Kleinberg in a network formation model that explicitly forces a power law property.

The small world phenomena suggests that in a connected social network, any two individuals are likely to be connected (i.e. know each other indirectly) by a short path. **Moreover, such a path can be found in a decentralized manner.**

In Milgram's 1967 small world experiment, he asked random people in Omaha Nebraska to forward a letter to a specified individual in a suburb of Boston which became the origin of the idea of six degrees of separation.

The six degrees of freedom phenomena

There are two basic ways for finding someone in a social network.

- We could ask all of our friends to tell all of their friends to tell all of their friends...(i.e. a traditional chain letter) that I am looking for person X.
- Now say assuming your online social network has a "broadcast to all" feature, this can be done easily but it has its drawbacks. Drawbacks?

The six degrees of freedom phenomena

There are two basic ways for finding someone in a social network.

- We could ask all of our friends to tell all of their friends to tell all of their friends...(i.e. a traditional chain letter) that I am looking for person X.
- Now say assuming your online social network has a "broadcast to all" feature, this can be done easily but it has its drawbacks. Drawbacks?
- Suppose that we want to reach someone and it either costs real money/effort to pass a message (e.g. postal mail) or perhaps I would prefer to not let everyone know that I am looking for person X. There is also possibly a "social cost" in terms of annoyance to people in the network receiving multiple requestss to pass on a message.

The six degrees of freedom phenomena

There are two basic ways for finding someone in a social network.

- We could ask all of our friends to tell all of their friends to tell all of their friends. . . (i.e. a traditional chain letter) that I am looking for person X.

- Now say assuming your online social network has a "broadcast to all" feature, this can be done easily but it has its drawbacks. Drawbacks?

- Suppose that we want to reach someone and it either costs real money/effort to pass a message (e.g. postal mail) or perhaps I would prefer to not let everyone know that I am looking for person X. There is also possibly a "social cost" in terms of annoyance to people in the network receiving multiple requestss to pass on a message.

- Clearly if everyone cooperates, the broadcast method ensures the shortest path to the intended target X in the leveled tree/graph of reachable nodes.

Reachable nodes without triadic closure

- If there is no triadic closure (i.e. your friends are not mutual friends, etc.), it is easy to see why every path is a shortest path to everyone in the network.
- Consider the number of people that you could reach by a path of length at most t if every person has say at least 5 friends.

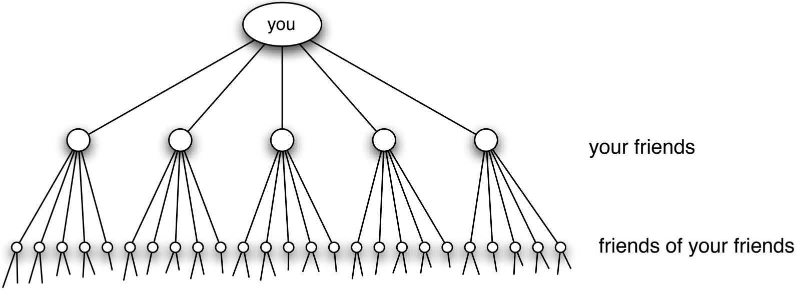


Figure: Pure exponential growth produces a small world [Fig 20.1 (a), E&K]

Reachable nodes with triadic closure

- Given that our friends tend to be mostly contained within a few small communities, the number of people reachable will be much smaller.

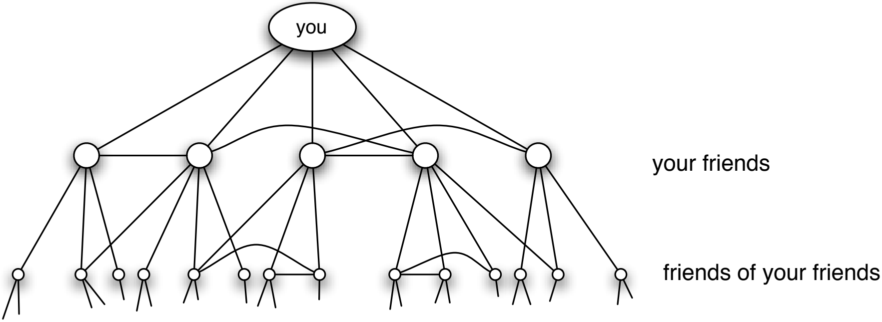


Figure: Triadic closure reduces the growth rate [Fig 20.1 (b), E&K]

The Watts-Strogatz model

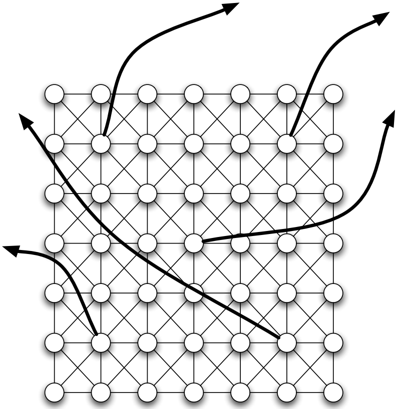
- Is it possible to have extensive triadic closure and still have short paths?

- Homophily (the tendency for people to seek out or be attracted to those who are similar to themselves) is consistent with triadic closure especially for strong ties whereas weak ties can connect different communities and thereby provide the kind of branching that yields short paths to many nodes.

- One stylized model to demonstrate the effect of these different kinds of ties is the Watts-Strogatz model, which considers nodes lying in a two dimensional grid and then having two types of edges:
  - Short-range edges to all nodes within some small distance r. This captures an idealized sense of homophily
  - A small number of random longer-distance edges to other nodes in the network; in fact, one needs very few such random edges to achieve the effect of short paths.

Very few random edges are needed

- A k by k “town” with probability $1/k$ that a person has a random weak tie.
- This would be sufficient to establish short paths.



[Fig 20.3, E&K]

But how does this explain the ability to find people in a decentralized manner

- In the Watts-Strogatz type of model, we can use the random edges (in addition to the short grid edges) and the geometric location of nodes to keep trying to reduce the grid distance to a target node.
  - This is analogous to the Milgram experiment where individuals seem to use geographic information to guide the search.
  - However, completely random edges does no reflect real social networks

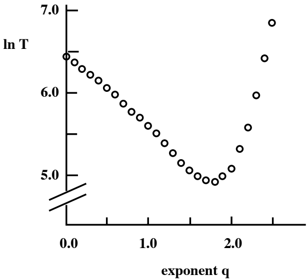
- Furthermore, having uniformly random edges will not work in general as:
  - Completely random edges (i.e. going to a random node anywhere in the network) are too random.
  - A random edge in an $n \times n$ grid is likely to have grid distance $\Theta(n)$.
  - Without some central guidance, such random edges will essentially just have us bounce around the network causing a substantially longer path to the target than the shortest path.

A modification of the model

- Random edges outside of ones "close community" are still more likely to reflect some relation to closeness.
- So assume as in the Watts-Strogatz model, from every node v we have edges to all nodes x within some grid distance r from v.
- And now, in addition, random edges are generated as follows: we (independently) create an edge from v to w with probability proportional to $d(v, w)^{-q}$ where $d(v, w)$ is the grid distance from v to w and $q \geq 0$ is called the clustering exponent.
- The smaller $q \geq 0$ is, the more completely random is the edge whereas large $q \geq 0$ leads to edges which are not sufficiently random and basically keeps edges within or very close to ones community.
- What is the best choice of $q \geq 0$?

So what is a good or the best choice of the clustering exponent q?

- It turns out that in this 2-dimensional grid model decentralized search works best when $q = 2$. (This is a result that holds and can be proven for the limiting behaviour, in the limit as the network size increases.)



[Fig 20.6, E&K]

  - Simulation of decentralized search in the grid-based model with clustering exponent q.
  - Each point is the average of 1000 runs on (a slight variant of) a grid with 400 million nodes.
  - The delivery time is best in the vicinity of exponent $q = 2$, as expected.
  - But even with this number of nodes, the delivery time is comparable over the range between 1.5 and 2.

More precise statements of Kleinberg's results on navigation in small worlds

The Milgram-like experiment

- Consider a grid network and construct (local contact) directed edges from each node u to all nodes v within grid distance $d(u, v) = k > 1$.
- Also probabilistically construct m (long distance) directed edges where each such edge is chosen with probability proportional to $d(v, w)^{-q}$ for $q \geq 0$.
- We think of k and m as constants and consider the impact of the clustering exponent q as the network size n increases.
- We assume that each node knows its location and the location of its adjacent edges and its distance to the location of a target node t.
- The Milgram-like experiment is that each node *tries* (without knowing the entire network) to move from a node u to a node v that is closest to t (in grid distance).