Strong and Weak Ties
Web Science (VU) (707.000)

Elisabeth Lex
KTI, TU Graz

April 20, 2015
1. Repetition

2. Strong and Weak Ties

3. General principles about structural properties of social networks

4. The Strength of Weak Ties

5. Graph Partitioning
Pivotal Nodes

- Node $k$ is pivotal for a pair of distinct nodes $i$ and $j$ if $k$ lies on every shortest path between $i$ and $j$
- In our example, node $B$ is pivotal for $(A,C)$ and $(A,D)$
- However, it is not pivotal for $(D,E)$

Some nodes are more important than the other nodes!
Node \( k \) is a **gatekeeper** if, for some other distinct nodes \( i \) and \( j \), \( k \) lies on every path between \( i \) and \( j \).

\[ \text{A is a gatekeeper because it lies on every path between B and E, or (D,E).} \]

**Degree** of a node is defined by the number of links connected to it - high degree means node is more central and thus more important.
Strong and Weak Ties
Motivation: Mark Granovetter

- Sociologist at Stanford University
- Best known for his theory “The strength of weak ties”
- Theory focuses on spread of information in social networks
- Strongly influenced by the Milgram experiment on social networks
(Interpersonal) Ties: Information carrying connections between people in a network, relationships between the nodes

3 types: strong, weak or absent

According to Granovetter, strength of ties is a combination of
- amount of time
- emotional intensity
- intimacy (mutual confiding)
- reciprocal services that characterize the tie

For a friendship network, link strength could represent frequency of contact, proximity, length of friendship, or subjective assessment of closeness

Strong ties are those with a larger weight (e.g. the friends you see more often or have known for longer)

Can you give examples of strong/weak ties?
Granovetter’s Experiment

Granovetter launched study on job seeking in the 1960ies and 1970ies to investigate usefulness of different types of ties in certain situations:

- How people find out about new jobs?
  - People find the information through personal contacts
  - But: contacts were often acquaintances (weak ties) rather than close friends (strong ties)!
  - Why acquaintances are most helpful?
General principles about structural properties of social networks
Strong Triadic Closure

- Property among three nodes A, B, C, such that if strong tie exists between A-B and A-C, there is a weak or strong tie between B-C.
- Too extreme for real-world networks, but good simplification.
- Useful to understand networks and predict network evolution.
- E.g. Trust networks: If A trusts B and A trusts C, B will probably trust C as well.

![Diagram](image)

**Figure:** Formation of E(B,C), since B and C they have common neighbor A.
Triadic Closure

- Occurs in social networks due to increased opportunity for two nodes with common neighbor to meet and to create at least a weak tie.

- Reasons for triadic closure:
  - **Opportunity**: B and C have a common friend A so there is an increased chance they will end up knowing each other.
  - **Trust**: B and C are friends with A. Gives them a basis for trusting each other that an arbitrary pair of unconnected people might lack.
  - **Incentive**: Node B has incentive to bring A and C together to decrease effort it takes to maintain two separate relationships.

- Networks that fulfill triadic closure are highly connected - thus high clustering coefficient.
  - Implication: low clustering coefficient implies few good friends.
General principles about structural properties of social networks

Applications for Triadic Closure Principle

- Can be used to model how networks will evolve
- Can predict development of ties in network
- Shows progression of connectivity
Measuring the Extent of Triadic Closure

- Repetition: CC of node A is defined as the probability that two randomly selected friends of A are friends with each other.
- CC for a node is a measure of the connectedness of its neighbourhood.
- Local CC of a node is a measure how transitive connections are in a network.
- I.e. a friend of a friend is also a friend.

\[
CC \text{ of } A = \frac{\text{Number of connections between friends of } A}{\text{Possible number of connections between friends of } A}
\]  

(1)
Repetition: Example for Clustering coefficient (CC)


\[
\text{CC of } A = \frac{\text{Number of connections between friends of } A}{\text{Possible number of connections between friends of } A} = \frac{1}{6} \quad (2)
\]

Figure: Clustering Coefficient computation for node A
Example Application for Strong and Weak Ties

What Are Lists?
A quick, optional way to organize friends so you can control what you see in your News Feed and post updates to specific people. We won’t tell your friends if you add them to these three new lists:

- **Close Friends**: Your best friends, who should show up more in News Feed.
- **Acquaintances**: Friends who should show up less in News Feed.
- **Restricted**: Friends who can only see posts and profile info you make public.

www.GROWUPGEEK.com

Elisabeth Lex  (KTI, TU Graz)  Networks  April 20, 2015  15 / 60
The Strength of Weak Ties
Usefulness of certain types of ties in certain situations

- Remember: Weak ties more helpful to find out about new jobs
- Granovetter: Weak ties can act as a “bridge” that spans parts of a social network
- Connect otherwise disconnected social groups
Bridges and Local Bridges

**Definition**

*A bridge between node A and node B exists if deleting the bridging edge would cause node A and node B to lie in two different components.*

**Figure:** Example of a bridge between the nodes A and B
Local Bridges (1/2)

- Bridges rare in real-world social networks (Small World Property)
- Local bridge occurs when node acts as gatekeeper between 2 neighbors who are otherwise not connected
- Local bridge: if nodes A and B have NO friends in common, i.e., if deleting the edge between A and B increases distance between A and B to value strictly more than 2

Figure: Edge E(A,B) is a local bridge since A and B have no common friends
Local Bridges (2/2)

Figure: Edge E(A,B) is a local bridge since A and B have no common friends.

Is there another local bridge in this example apart from E(A,B)?
Local Bridges (2/2)

Figure: Edge E(A,B) is a local bridge since A and B have no common friends.

Is there another local bridge in this example apart from E(A,B)?

No - Local bridges never form the side of any triangle in the network!
Span of Local Bridges

- Span of a local bridge is the distance btw its endpoints if the edge is deleted
- Length of the shortest path between two nodes
- Local bridges with large span play similar role as bridges since their endpoints provide access to parts of network that would otherwise be far away
Span of Local Bridges

What is the span of the local bridge $E(A,B)$?
Span of Local Bridges

What is the span of the local bridge $E(A,B)$? 4
Remember Granovetter’s experiment

- Why are acquaintances more important?
Remember Granovetter’s experiment

Why are acquaintances more important?

A, C, D, and E will all tend to be exposed to similar sources of info, while A’s link to B offers access to things A otherwise wouldn’t necessarily hear about!
Local bridges and weak ties

The relationship between local bridges and weak ties through the strong triadic closure is

- if a node A satisfies **strong triadic closure** and is **involved in at least two strong ties** then any local bridge adjacent to A **must be a weak tie**
Local bridges and weak ties

Proof by contradiction:

- Node A satisfies strong triadic closure and is involved in at least 2 strong ties
- Suppose $E(A,B)$ is a local bridge
- Strong Triadic Closure says: $E(B,C)$ must exist
- But since $E(A,B)$ is a bridge, $E(B,C)$ must NOT exist since endpoints of a bridge have no friends in common
Summary: The Strength of Weak Ties

- Weak ties (acquaintances) are social ties that connect us to new sources of information
- Weak ties link together tightly knit communities, each containing a large number of strong ties
- Local bridge property directly related to their weakness as social ties
- Dual role of being weak connections but also valuable links to parts of the network that are harder to access - is the strength of weak ties
Example: Weak Ties

- Example: Spread of information/rumors in social networks
- Studies have shown that people rarely act on mass-media information unless it is also transmitted through personal ties [Granovetter 2003, p 1274]
- Information/rumors moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones, bridges will not be crossed
**Summary**

- **Bridges**: edges when removed will find their endpoints in different connected components.
- **Local bridges**: edges whose endpoints are not in triangles.
- **2 types of edges**: weak and strong ties.
- **Strong triadic closure**: Two strong ties imply a third strong / weak tie.
- **Local bridges are weak ties**.
Real world validation: Onnela et al., 2007 (1/5)

- Investigated a large cellphone network with 4 millions users
- Edge between two users if they called each other within the observation time period of 18 months
- 84% of nodes in a giant component, i.e. a single connected component containing most of the individuals in the network
- Tie strength: time spent in a phone conversation
Extended definition of local bridge to **neighbourhood overlap** to detect “almost” local bridges

Since only very small fraction of edges in the cell-phone data were local bridges

\[
\text{Neighbourhood Overlap} = \frac{\# \text{ nodes who are neighbours of both } A \text{ and } B}{\# \text{ nodes who are neighbours of at least } A \text{ or } B}
\]

\[0 \text{ if } E(A,B) \text{ is a local bridge}
\]

Note that neighborhood overlap increases with increasing tie strength!
Real world validation: Onnela et al., 2007 (3/5)

E.g. Neighborhood Overlap of nodes A-F:
- Denominator determined by B, C, D, E, G, J
- Nominator: only C is a neighbor of both A and F
- Result: Neighborhood overlap: 1/6
Real world validation: Onnela et al., 2007 (4/5)

Figure: Neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength

- Shows relation between tie strength and neighborhood overlap
- As tie strength decreases, overlap decreases, i.e. weak ties becoming almost local bridges
To support the hypothesis that weak ties tend to link together more tightly knit communities, Onnela et al. perform two simulations:

- Removing edges in decreasing order of tie strength, the giant component shrank gradually.
- Removing edges in increasing order of tie strength, the giant component shrank more rapidly and at some point then started fragmenting into several components.

Be careful when investigating networks only in respect to structural properties - because one knows relatively little about the meaning or significance of any particular node or edge!
Marlow et al. (2009) investigated 4 types of Facebook networks:

- All friends
- Maintained relationships (passive, e.g. following a user)
- One way communication of a user (e.g. liking content the other user posted)
- Reciprocal (mutual) communication (e.g. chatting on the wall)
Strong and weak ties in large social networks: Facebook (2/4)

They found that

- All networks thin out when links represent stronger ties.
- As the number of total friends increases, the number of reciprocal communication links levels out at slightly more than 10
Strong and weak ties in large social networks: Facebook (3/4)

Figure: 4 types of networks on Facebook and their graphs
Strong and weak ties in large social networks: Facebook (4/4)

Figure: 4 types of networks on FB as function of total neighborhood size for users.
Huberman et al. (2009) investigated weak ties on Twitter:

- **Dataset:** Followers graph (directed)
- **Strong ties:** users to whom the user sent at least 2 messages in the observation period
- **Weak ties:** users followed
Figure: Total number of user’s strong ties as function of number of followees on Twitter
Graph Partitioning
Weak ties and communities

Figure: Co-authorship network of physicists and applied mathematicians.

- Weak ties seem to link together groups of tightly-knit nodes, each group containing a large number of strong ties.
- How can we find these groups, aka communities?
Why are communities interesting?

For example to:

- Detect clusters of customers with similar interests and use them to personalize recommendation systems
- Study relationships among nodes to detect nodes that connect communities and thus act as information brokers
- Identification of clusters in networks can be exploited e.g. to improve PageRank
- ...
The Notion of Betweenness

Betweenness:

- A way to measure importance of a node or an edge in a network
- (Remember degree as a measure of importance - the more neighbours, the more important the node)
- **Betweenness centrality of a node** $A$ is the sum of the fraction of all-pairs shortest paths that pass through $A$
- Which means, nodes where a lot of transit can happen are important
- **Betweenness centrality of an edge** $E(A,B)$ is the sum of the fraction of all-pairs shortest paths that pass through $E(A,B)$
- i.e. the number of shortest paths that pass through a given edge
**Example**

*Figure:* Nodes in “Les Miserables” network: betweenness centrality

- Node “Valjean” has highest betweenness centrality, sits on many shortest paths
- E.g. Fantine has tightly-knit local communities connected to it
- Thickness of edges represent tie strength (e.g. Valjean’s strongest ties are with Marius and Cosette)
Motivation

- Bridges and local bridges often connect densely connected parts of the network
- Like important highway parts, local bridges are important because without them, paths between pair of nodes need to be re-routed a longer distance
- Good candidates for removal in graph partitioning as these edges carry most traffic ("flow") on the network
- Goal: find edges that transport most flow over the network!
Challenges when computing betweenness

- According to definition: consider all the shortest paths between all pairs of nodes ("sum of the fraction of all-pairs shortest paths that pass through E(A,B)")
- Computationally intensive!
- One solution: Method based on Breadth-first Search (BFS)
Repetition: Breadth-first search

- Begin at a given node i in the network
- All neighbors of i (nodes connected to i) are at distance 1
- Then find all neighbors of these neighbors (not counting nodes that are already neighbors of i) and they are distance 2
- Then find all neighbors of the nodes from the previous step (not counting nodes already found at distance 1 and 2) and they are at distance 3
- Continue and search in next layers at next distance until no new nodes can be found
Consider graph from perspective of one node at a time; for each given node, compute how the total flow from that node to all others is distributed over the edges. Do this for every node, and then simply add up the flows from all of them to get the betweennesses on every edge.

For each node A:

- BFS starting at A
- Determine number of shortest paths from A to every other node
- Based on these numbers, determine the amount of “flow” from A to all other nodes that use each edge
Example

(a) A sample network

(b) Breadth-first search starting at node A
Graph Partitioning

Example

- Each node in the first layer has only 1 shortest path from A
- The number of shortest paths to each other node is the sum of the number of shortest paths to all nodes directly above it
- Avoid finding the shortest paths themselves
Example: Determining Flow Values

- Start with node K: Single unit of flow arrives at K + equal number of shortest paths from A to K through nodes I and J. So, flow equally divided over two incoming edges (1/2 on both edges).
- Total amount of flow arriving at I is equal to the one unit actually destined for I plus the half-unit passing through to K, for a total of 3/2. 2 shortest paths from A through F, 1 through G - thus, 1 on F, 1/2 on G.
- Continue upwards until finished
- Finally, sum up flow values to get betweenness value for each edge.
Summing Up

- Build one BFS structures for each node
- Determine flow values for each edge using the described procedure
- Sum up flow values of each edge in all BFS structures to get its betweenness value
- Since flow between each pair of nodes X and Y is counted twice (once when BFS from X and once when BFS from Y), at the end, divide everything by two
- Use these betweenness values to identify edges of highest betweenness for purposes of removing them using the Girvan-Newman method described next.
Girvan-Newman Method

Based on successively deleting edges of high betweenness as they are the most important edges for connecting networks

1. Find the edge of highest betweenness — or multiple edges of highest betweenness, if there is a tie — and remove these edges from the graph. This may cause the graph to separate into multiple components. If so, this is the first level of regions in the partitioning of the graph.

2. Now recalculate all betweennesses, and again remove the edge or edges of highest betweenness. This may break some of the existing components into smaller components; if so, these are regions nested within the larger regions.

(...). Proceed in this way as long as edges remain in graph, in each step recalculating all betweennesses and removing the edge or edges of highest betweenness.
Example

(a) Step 1

(b) Step 2

(c) Step 3
Girvan-Newman Method

- Algorithm is divisive and recursive
- Tries to find an optimal way of cutting the graph into two pieces, and then it does the same on the pieces.
- Each time it finds a way to split a graph into two optimal pieces, it calls itself on the pieces.
- Terminates when it is given a graph with only one node.
- Returns list of all the components it has found
Summary

We have learned about:

- Tie strength: Strong and weak ties
- Strong triadic closure
- Bridges and local bridges
- The strength of weak ties
- Weak ties and communities - edge betweenness (Girvan Newman Method)
Thanks for your attention - Questions?

elisabeth.lex@tugraz.at

Slides use figures from Chapter 3 of Networks, Crowds and Markets by Easley and Kleinberg (2010)