Title: groh: profile1 (06.05.2015)
Date: Wed May 06 08:13:57 CEST 2015
Duration: 91:37 min
Pages: 55

**Social Network**

Slightly refined Social Network Model: Graph $G = (V, E, P_V, P_E, f_V, f_E)$.  
- **Nodes** $V = \bigcup V_i$: represent humans (actors) of “sorts” ($\leftrightarrow$ modes) $V_i$.
- **Edges** $E \subseteq V \times V$; $E = \bigcup E_i$: represent directed binary social relations (ties) of “sorts” $E_i$.  
  - $P_V$: Set of Node Profiles  
  - $P_E$: Set of Edge Profiles  
  - $f_V: V \rightarrow P_V$  
  - $f_E: V \rightarrow P_E$  

**Social Network**

Slightly refined Social Network Model: Graph $G = (V, E, P_V, P_E, f_V, f_E)$.  
- **Nodes** $V = \bigcup V_i$: represent humans (actors) of “sorts” ($\leftrightarrow$ modes) $V_i$.
- **Edges** $E \subseteq V \times V$; $E = \bigcup E_i$: represent directed binary social relations (ties) of “sorts” $E_i$.  
  - $P_V$: Set of Node Profiles  
  - $P_E$: Set of Edge Profiles  
  - $f_V: V \rightarrow P_V$  
  - $f_E: V \rightarrow P_E$
6 degrees of separation

av. path length in real world SN is \( \sim 6 \)

- First occurrence of a claim similar to 6 degrees: G. Marconi (Italian Physicist & Nobel Prize laureate) 1909: Number of radio stations necc. to cover inhabited world \( \rightarrow \) any transmission path needs about 6 stations

- 1920s: Hungarian writer F. Karinthy claims six degrees of separation in Budapest in a short story (prob. inspired by Marconi)

- Most famous: S. Milgram (Inspired by unpublished paper by M. Kochen & I. de Sola Pool claiming \( \sim 3 \) degrees in USA): “Small world experiment” [20]. Randomly chosen people: mail letter to target person; track record \( \rightarrow \sim 6 \) av. path length

6 degrees of separation

- Popular culture: Erdős number / Kevin-Bacon Number / Erdős-Bacon-Number

- Several newer experiments (see [20], [21]) on degree of separation on the web (Facebook, Email-studies (D. Watts, Columbia U.) etc.) also showed degree of separation \( \sim 6 \)

- More thorough mathematical investigation \( \rightarrow \) Random Graph Theory

- Watts and Strogatz [22]: Small World graph (informal): high clustering coefficient, small mean av. Path-length

Technical intermezzo: Clustering coefficient

- Undirected Graph: Clustering Coefficient \( C_i \) of node \( v_i \): Measures “how close” \( v_i \) and its neighbors \( \{ v_j \in N_i \} \) (where neighborhood \( N_i \) is \( \{ v_j | \{ v_i, v_j \} \in E; E \subset (V \bigcup \{ v \}) \} \) ) are to a complete subgraph (clique):

\[
C_i = \frac{\{ e_{ij} | v_i, v_j \in N \}}{d_i(d_i-1)}
\]

- Directed Graph: Clustering Coefficient \( C_i \) of node \( v_i \): Measures “how close” \( v_i \)’s neighbors \( \{ v_j \in N_i = N_i^{out} \bigcup N_i^{in} \} \) (where out-neighborhood \( N_i^{out} \) is \( \{ v_j | (v_i, v_j) \in E; E \subset (V \bigcup V) \} \) and in-neighborhood \( N_i^{in} \) is \( \{ v_j | (v_j, v_i) \in E; E \subset (V \bigcup V) \} \) ) are to a complete subgraph (clique):

\[
C_i = \frac{\{ e_{ij} | v_i, v_j \in N \}}{d_i(d_i-1)}
\]
**Technical intermezzo: Clustering coefficient**

- **Undirected Graph: Clustering Coefficient** $C_i$ of node $v_i$: Measures "how close" $v_i$ and its neighbors $\{v_j \in N_i\}$ (where neighborhood $N_i$ is $\{v_j | \{v_i, v_j\} \in E; E \subseteq (V\times V)\}$) are to a complete subgraph (clique):

  $$C_i = \frac{|\{e_{ij} | v_i, v_j \in N_i\}|}{\binom{d_i}{2}}$$

  Degree $d_i$ of node $v_i$:

  $$d_i = |N_i|$$

- **Directed Graph: Clustering Coefficient** $C_i$ of node $v_i$: Measures "how close" $v_i$'s neighbors $\{v_j \in N_i^o \cup N_i^i\}$ (where out-neighborhood $N_i^o$ is $\{v_j | \{v_i, v_j\} \in E; E \subseteq (V \times V)\}$ and in-neighborhood $N_i^i$ is $\{v_j | \{v_j, v_i\} \in E; E \subseteq V \times V\}$) are to a complete subgraph (clique):

  $$C_i = \frac{|\{e_{ij} | v_i, v_j \in N_i\}|}{\binom{d_i}{2}}$$

  Degree $d_i$ of node $v_i$:

  $$d_i = |N_i|$$

---

**History of Social Network Analysis, Main Contributors**

- See e.g. [9]:

  - Main contributing fields of science: Sociology (surprisingly ☃️), Anthropology, Urban Studies, Mathematics (modeling & evaluation formalisms), Physics (large community (surprisingly)), Computer Science (graph algorithms etc.), Economic Sciences

  - 1887: F. Tönnies (German sociologist): 2 basic "sorts" of groups: Gemeinschaft (Family, Friends etc.; supported by "Wesenswillen") ↔ Gesellschaft (Goal oriented; Film, State etc.; supported by "Kührwille")

  - 1911: G. Simmel (German sociologist): Sociability of humans (especially in larger cities): One of the first to impose a "social network" view


History of Social Network Analysis, Main Contributors

- 2000s-present: A. Barabasi, D. Watts, M. Newman, J. Kleinberg: (Physicists take over*), A. Pentland (Reality Mining) etc.

History of Social Network Analysis, Main Contributors

- 2000s-present: A. Barabasi, D. Watts, M. Newman, J. Kleinberg: (Physicists take over*), A. Pentland (Reality Mining) etc.
Centrality

- Centrality indices formalize intuitive feeling that some nodes (or edges) are more central (important, meaningful etc.) than others.

- Interpretations of “centrality”: “influence”, “prestige”, “control”, “heavily required for information flow”

- Example: n persons vote for a leader; (u,v) ∈ E if \( u \) voted for \( v \); Winner (most central node): node with most incoming edges (highest in-degree).

  → Degree Centrality

  Other variant: (u,v) ∈ E if \( u \) has convinced \( v \) to vote for \( u \)’s favorite candidate. (Influence network) → node with large out-degree is central

- Other Example: If graph can be split up into groups X and Y and if node \( u \) has many edges to \( X \) and many edges to \( Y \) → \( u \) mediates most information between groups → \( u \) is central

  → Betweenness centrality

General “Definition”: Structural Index

- “Importance” has many aspects but minimal def. for centrality: Only depends on structure of graph:

- Structural Index: Let \( G = (V,E,w) \) be a weighted directed or undirected multigraph. A function \( s: V \rightarrow \mathbb{R} \) (or \( s: E \rightarrow \mathbb{R} \)) is a structural index iff

  \[ \forall x: G \cong H \rightarrow s_G(x) = s_H(\phi(x)) \]

  (Recall: Two graphs \( G \) and \( H \) are isomorphic (\( G=H \)) iff exists a bijective mapping \( \Phi: G \rightarrow H \) so that \( (u,v) \in G \iff (\Phi(u),\Phi(v)) \in H \))

- structural index induces order (\( \leq \)) on nodes/edges

- centrality can usually only be viewed as measured on an ordinal scale only (not interval or ratio scale)

Distance- and Neighborhood-based Centralities

- Centrality measures defined on the basis of distances or neighbourhoods in the graph:

  Centrality of vertex \( \leftrightarrow \) “reachability” of a vertex

Neighborhoods: Degree Centrality

- Most basic: Degree centrality: \( c(u) = \text{deg}(u) \) (or \( c(u)=\text{in-deg}(u) \) or \( c(u) = \text{out-deg}(u) \)) → local measure

- Applicable: If edges have “direct vote” semantics
Distances: Eccentricity

- **Example:** Facility location problems: Objective function on $d(u,v)$; e.g. minimax (minimize maximal distance (e.g.: hospital emergency)) $\Rightarrow$ can be mapped to social case

- For the moment: $G$ is **undirected and unweighted** (e.g. “friendship”). Mapping to weighted and/or directed case is possible.

- **Eccentricity** $e(u) = \max\{d(u,v); v \in V\}$

- **Eccentricity** $e(u) = \max\{d(u,v); v \in V\}$

- **Center of a graph:** Set of all nodes with minimum eccentricity

- **Eccentricity based centrality measure:**
  
  $$c(u) = \frac{1}{e(u)} = \frac{1}{\max\{d(u,v); v \in V\}}$$

- **→ nodes in the center of the graph have maximal centrality**

**Diagram:**

- Graph with $e(u)$ values

- Graph with $e(u)$ values
**Distances: Closeness**

- **Minimum problem**: find nodes whose sum of distances to other nodes is minimal (service facility location problem). For all $u$ minimize total sum of minimal distances $\sum_{v \neq u} d(u, v)$

- **Social analog**: Determine central figure for coordination tasks

- **Example**:

\[
\begin{array}{cccccc}
36 & 26 & 24 & 22 & 32 \\
\end{array}
\]

graph with $\sum_{v \neq u} d(u, v)$ values

**Distances: Closeness**

- **Possible resulting centrality index: closeness**:
  \[
c(u) = \sum_{v \in V} \frac{1}{d(u, v)}
\]
  Only applicable to connected graphs; disconnected graph: all nodes will get the same centrality $1/|V|

- **Other possibility**
  \[
c(u) = \sum_{v \in V} \frac{(\Delta_G + 1 - d(u, v))}{|V| - 1}
\]
  $\Delta_G$ is the diameter of the graph

- **If computed on directed graph**: (set $d(u, u) = 0$ and set $d(u, v) = 0$ if $u, v$ are unreachable via directed path → problematic!); using in-distances: "integration", using out-distances "radiality" (see [6])

**Distances: Closeness**

- **Possible resulting centrality index: closeness**:
  \[
c(u) = \sum_{v \in V} \frac{1}{d(u, v)}
\]
  Only applicable to connected graphs; disconnected graph: all nodes will get the same centrality $1/|V|

- **Other possibility**
  \[
c(u) = \sum_{v \in V} \frac{(\Delta_G + 1 - d(u, v))}{|V| - 1}
\]
  $\Delta_G$ is the diameter of the graph

- **If computed on directed graph**: (set $d(u, u) = 0$ and set $d(u, v) = 0$ if $u, v$ are unreachable via directed path → problematic!); using in-distances: "integration", using out-distances "radiality" (see [6])
Distances: Closeness

- Possible resulting centrality index: closeness:
  
  \[ c(u) = \frac{1}{\sum_{v \in V} d(u, v)} \]
  
  Only applicable to connected graphs; disconnected graph: all nodes will get the same centrality \( 1/\infty \)

- Other possibility
  
  \[ c(u) = \frac{\sum_{v \in V} (\Delta_G + 1 - d(u, v))}{|V| - 1} \]
  
  \( \Delta_G \) is the diameter of the graph
  
  If computed on directed graph: (set \( d(u, v) = 0 \) and set \( d(v, u) = 0 \) if \( u, v \) are unreachable via directed path → problematic !); using in-distances: „integration“, using out-distances „radiality“ (see [6])

Distances: Centroids

- Possible resulting centrality index: closeness:
  
  \[ c(u) = \frac{1}{\sum_{v \in V} d(u, v)} \]
  
  Only applicable to connected graphs; disconnected graph: all nodes will get the same centrality \( 1/\infty \)

- Other possibility
  
  \[ c(u) = \frac{\sum_{v \in V} (\Delta_G + 1 - d(u, v))}{|V| - 1} \]
  
  \( \Delta_G \) is the diameter of the graph
  
  If computed on directed graph: (set \( d(u, v) = 0 \) and set \( d(v, u) = 0 \) if \( u, v \) are unreachable via directed path → problematic !); using in-distances: „integration“, using out-distances „radiality“ (see [6])

- Competitive objective: Given number of competitors: where to open a store (Customers will just choose store based on minimal distance)?

- Social Problem: Example: find “social ecological niche”

- Formalization: For \( u, v : \gamma_u(v) \) = number of vertices closer to \( u \) than to \( v \); if one salesman selects \( u \) and competitor selects \( v \) as locations, the first will have
  
  \[ \gamma_u(v) = \frac{1}{2} (|V| - \gamma_u(v) - \gamma_v(u)) = \frac{1}{2} |V| + \frac{1}{2} (\gamma_u(v) - \gamma_v(u)) \]
  
  customers
**Competitive objective:** Given number of competitors: where to open a store (Customers will just choose store based on minimal distance)?

**Social Problem:** Example: find "social ecological niche"

**Formalization:** For $u, v : \gamma_u(v)=$number of vertices closer to $u$ than to $v$; If one salesman selects $u$ and competitor selects $v$ as locations, the first will have

$$\gamma_u(v) + \frac{1}{2}(|V| - \gamma_u(v) - \gamma_u(u)) = \frac{1}{2} |V| + \frac{1}{2} (\gamma_u(v) - \gamma_u(u))$$

customers

**Competitive objective:** Given number of competitors: where to open a store (Customers will just choose store based on minimal distance)?

**Social Problem:** Example: find "social ecological niche"

**Formalization:** For $u, v : \gamma_u(v)=$number of vertices closer to $u$ than to $v$; If one salesman selects $u$ and competitor selects $v$ as locations, the first will have

$$\gamma_u(v) + \frac{1}{2}(|V| - \gamma_u(v) - \gamma_u(u)) = \frac{1}{2} |V| + \frac{1}{2} (\gamma_u(v) - \gamma_u(u))$$

customers

**→** Competitor will want to minimize

$$f(u, v) = \gamma_u(v) - \gamma_u(u)$$

**→** Possible centrality index: First salesman knows the strategy of the competitor and calculates for each location the worst case:

$$c(u) = \min_v \{ f(u, v) : v \in V \setminus \{u\} \}$$

**c(u) is called centroid value:** measures the advantage of location $u$ compared to other locations: Minimal loss of customers if he choses $u$ and a competitor choses $v$
Distances: Centroids

- Competitor will want to minimize
  \[ f(u, v) = \gamma_u(v) - \gamma_v(u) \]

- **Possible centrality index**: First salesman knows the strategy of the competitor and calculates for each location the worst case:
  \[ c(u) = \min_v \{ f(u, v) : v \in V \setminus \{u\} \} \]

- \( c(u) \) is called centroid value: measures the advantage of location \( u \) compared to other locations: Minimal loss of customers if he chooses \( u \) and a competitor chooses \( v \).

---

Shortest Paths

- Indices of this section can be applied to weighted, unweighted, directed, undirected and multigraphs and to edges and vertices ("graph elements" \( \chi \)).

- Assume that set of all shortest paths APSP is known (e.g. by application of Floyd Warshall algorithm in \( O(|V|^3) \) worst case time).

**Reminder:**
- BFS: SSSP; \( O(|V|+|E|) \) worst case time complexity, edge-weights\( = \) 1
- Dijkstra: SSSP; \( O(|V| \log |V| +|E|) \) with Fibonacci heap; edge-weights\( \geq 0 \)
- Floyd Warshall: APSP; \( O(|V|^3) \) worst case time, arbitrary weights, no negative cycles allowed (but can be detected via the alg.), dynamic programming;
- Bellman Ford: SSSP; \( O(|V| |E|) \), arbitrary weights, no negative cycles allowed (but can be detected via the alg.).
Shortest Paths

- Indices of this section can be applied to weighted, unweighted, directed, undirected and multigraphs and to edges and vertices (“graph elements” $x$).
- Assume that set of all shortest paths APSP is known (e.g. by application of Floyd Warshall algorithm in $O(|V|^3)$ worst case time)

Reminder:
- BFS: SSSP; $O(|V|+|E|)$ worst case time complexity, edge-weights $==1$
- Dijkstra: SSSP; $O(|V| \log |V| + |E|)$ with Fibonacci heap; edge-weights $\geq 0$
- Floyd Warshall: APSP, $O(|V|^3)$ worst case time, arbitrary weights, no negative cycles allowed (but can be detected via the alg.), dynamic programming;
- Bellman Ford: SSSP; $O(|V||E|)$, arbitrary weights, no negative cycles allowed (but can be detected via the alg.)

Shortest Paths: Stress

- Heuristic: If a vertex is part of many shortest paths $\rightarrow$ “much information will run through it” if information is routed along shortest paths
- Social analog: People that are asked to contribute to a workflow more often than others
- $\rightarrow$ A vertex $v$ is more central the more shortest paths run through it. Let $\sigma_{ab}(v)$ denote the number of shortest paths from node $a$ to node $b$ containing $v$. $\sigma_{ab}(v)$ can be $>1$ if there are several paths with the same minimal length

stress centrality: $c(v) = \sum_{a \neq v} \sum_{b \neq v} \sigma_{ab}(v)$
Shortest Paths: Stress

- Variant for edges:

\[
c'(e) = \sum_{a \in V} \sum_{b \in V} \sigma_{ab}(e)
\]

Shortest Paths: Shortest Path Betweenness

- Again assume that communication (workflows etc.) happen along shortest paths only. Let

\[
\delta_{ab}(v) = \frac{\sigma_{ab}(v)}{\sigma_{ab}}
\]

with \(\sigma_{ab}\): total number of shortest paths between nodes a and b.

**Interpretation.** Probability that \(v\) is involved in a communication between a and b.
**Shortest Paths: Stress**

- Variant for edges:
  \[ c'(e) = \sum_{a \in V} \sum_{b \in V} \sigma_{ab}(e) \]

---

**Shortest Paths: Shortest Path Betweenness**

- Shortest Path Betweenness (SPB) centrality is then:
  \[ c(v) = \sum_{a \in V} \sum_{b \in V} \delta_{ab}(v) \]

  - Interpretation: Control that \( v \) exerts on the communication in the graph
  
  - Also applicable to disconnected graphs
  
  - Algorithm by Ulrik Brandes computes SPB in \( O(|V||E| + |V|^2 \log |V|) \) time

---

**Shortest Paths: Shortest Path Betweenness**

- Shortest Path Betweenness (SPB) centrality is then:
  \[ c(v) = \sum_{a \in V} \sum_{b \in V} \delta_{ab}(v) \]

  - Interpretation: Control that \( v \) exerts on the communication in the graph

  - Also applicable to disconnected graphs

  - Algorithm by Ulrik Brandes computes SPB in \( O(|V||E| + |V|^2 \log |V|) \) time
**Example** why shortest path betweenness centrality (now denoted as \(c_{SPB}\)) might be more interesting than the basic stress centrality (now denoted as \(c_S\)):

Each node \(\bullet\) has:
- \(c_S = 28\)
- \(c_{SPB} = 1/3 \times 28\)

**Shortest Path Betweenness (SPB) Centrality** is then:

\[
c(v) = \sum_{a \in V} \sum_{b \in V} \delta_{ab}(v)
\]

**Interpretation:** Control that \(v\) exerts on the communication in the graph.

Also applicable to disconnected graphs.

Algorithm by Ulrik Brandes computes SPB in \(O(|V||E| + |V|^2 \log |V|)\) time.