Script generated by TTT

Title: profile1 (04.06.2013)

Date: Tue Jun 04 12:02:22 CEST 2013

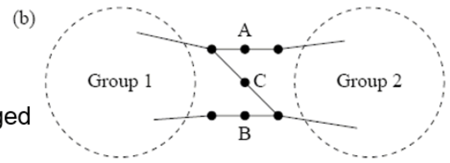
Duration: 89:34 min

Pages: 52

Critique on Betweenness Based Centralities

- major critique: Max-Flow betweenness centrality (suggested to counteract this drawback) may exhibit similar problems

- here: special Max-Flow betweenness centrality mfb:
  - -- limit edge capacity to one
  - -- mfb(i) := maximum possible flow through i over all possible solutions to the s-t-maximum flow problem, averaged over all s and t.



(b) In calculations of flow betweenness, vertices A and B in this configuration will get high scores while vertex C will not.

Source: [5]

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- cfb == random walk betweenness centrality (rwb):

- rwb(i): move around „messages“: start (absorbing) random walk at s, end at t:

  rwb(i):= net number of times that a message passes through i on its journey (averaged over a large number of trials and averaged over s, t)

  („net“ number of times: „cancel back and fourth passes“)

- if in i, probability that in next step j:

$$M_{ij} = \frac{A_{ij}}{k_j}, \qquad \text{for } j \neq t,$$

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-1} \qquad \text{with } D = \text{diag}(k_i)$$

$$D_{ii} = k_i$$

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- cfb == random walk betweenness centrality (rwb):

- rwb(i): move around „messages“: start (absorbing) random walk at s, end at t:

  rwb(i):= net number of times that a message passes through i on its journey (averaged over a large number of trials and averaged over s, t)

  („net“ number of times: „cancel back and fourth passes“)

- if in i, probability that in next step j:

$$M_{ij} = \frac{A_{ij}}{k_j}, \qquad \text{for } j \neq t,$$

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-1} \qquad \text{with } D = \text{diag}(k_i)$$

$$D_{ii} = k_i$$

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- cfb == random walk betweenness centrality (rwb):
- rwb(i): move around „messages“: start (absorbing) random walk at s, end at t:

  rwb(i):= net number of times that a message passes through i on its journey (averaged over a large number of trials and averaged over s, t)

  („net“ number of times: „cancel back and fourth passes“)
- if in i, probability that in next step j:

$$M_{ij} = \frac{A_{ij}}{k_j}, \quad \text{for } j \neq t,$$

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-1}$$ with D = diag(k_i)

$$D_{ii} = k_i$$

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- we never leave t, once we get there (“Hotel California effect” :-)) →

$$M_{it} = 0 \text{ for all } i$$

→ possible: remove column t without affecting transitions between any other vertices;

denote by $\mathbf{M}_t = \mathbf{A}_t \cdot \mathbf{D}_t^{-1}$ the matrix with these elements removed, and similarly for A_t and D_t.

- for a walk starting at s, the probability that we find ourselves at vertex j after r steps is given by $[\mathbf{M}_t^r]_{js}$
- probability that we then take a step to an adjacent vertex i is

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- we never leave t, once we get there (“Hotel California effect” :-)) →

$$M_{it} = 0 \text{ for all } i$$

→ possible: remove column t without affecting transitions between any other vertices;

denote by $\mathbf{M}_t = \mathbf{A}_t \cdot \mathbf{D}_t^{-1}$ the matrix with these elements removed, and similarly for A_t and D_t.

- for a walk starting at s, the probability that we find ourselves at vertex j after r steps is given by $[\mathbf{M}_t^r]_{js}$
- probability that we then take a step to an adjacent vertex i is

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

$$\sum_{r=0}^{\infty} M^r = (I - M)^{-1}$$ if $\forall i: |\lambda_i| < 1$ where λ_i Eigenvalues of M

$$k_j^{-1}[(\mathbf{I}-\mathbf{M}_t)^{-1}]_{js}.$$

$$\rightarrow \quad \mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I} - \mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;

net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

$$\sum_{r=0}^{\infty} M^r = (I - M)^{-1} \quad \text{if} \quad \forall i: |\lambda_i| < 1 \quad \text{where } \lambda_i \text{ Eigenvalues of M}$$

→ $$\mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I} - \mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;
net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

$$\sum_{r=0}^{\infty} M^r = (I - M)^{-1} \quad \text{if} \quad \forall i: |\lambda_i| < 1 \quad \text{where } \lambda_i \text{ Eigenvalues of M}$$

→ $$\mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I} - \mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;
net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

  the total number of times V_{ji} we go from j to i, averaged over all possible walks is

$$k_j^{-1}[(\mathbf{I} - \mathbf{M}_t)^{-1}]_{js}.$$

→ $$\mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I} - \mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;
net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

  the total number of times V_{ji} we go from j to i, averaged over all possible walks is

$$k_j^{-1}[(\mathbf{I} - \mathbf{M}_t)^{-1}]_{js}.$$

→ $$\mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I} - \mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;
net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- previous slide: probability at j after r steps and then j → i was:

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

- summing over r from 0 to ∞ : → geometric series →

  the total number of times V_{ji} we go from j to i, averaged over all possible walks is

$$k_j^{-1}[(\mathbf{I}-\mathbf{M}_t)^{-1}]_{js}.$$

$$\rightarrow \qquad \mathbf{V} = \mathbf{D}_t^{-1}\cdot(\mathbf{I}-\mathbf{M}_t)^{-1}\cdot\mathbf{s} = (\mathbf{D}_t-\mathbf{A}_t)^{-1}\cdot\mathbf{s}.$$

as before: the net flow of the random walk along the edge from j to i == $|V_i - V_j|$;
net flow through vertex i is a half the sum of the flows on the incident edges

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- current flow at node i:

$$I_i^{(st)} = \frac{1}{2}\sum_j A_{ij}|V_i^{(st)} - V_j^{(st)}|$$
$$= \frac{1}{2}\sum_j A_{ij}|T_{is} - T_{it} - T_{js} + T_{jt}|, \quad \text{for } i \neq s,t.$$

- unit current flow at nodes s and t:

$$I_s^{(st)} = 1, \qquad I_t^{(st)} = 1.$$

- cfb(i) (denoted as b_i) is then:

$$b_i = \frac{\sum_{s<t} I_i^{(st)}}{\frac{1}{2}n(n-1)}.$$

(takes O(m n^2) for all i) →
(plus matrix inversion:)
O((m+n) n^2) for everything

Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- we never leave t, once we get there ("Hotel California effect" :-)) →

$$M_{it} = 0 \text{ for all } i$$

→ possible: remove column t without affecting transitions between any other vertices;

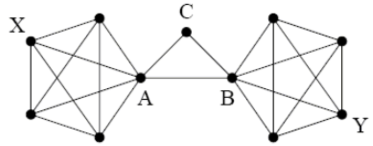
denote by $\mathbf{M}_t = \mathbf{A}_t \cdot \mathbf{D}_t^{-1}$ the matrix with these elements removed, and similarly for A_t and D_t.

- for a walk starting at s, the probability that we find ourselves at vertex j after r steps is given by $[\mathbf{M}_t^r]_{js}$

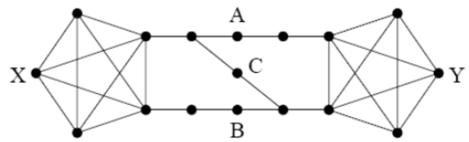
- probability that we then take a step to an adjacent vertex i is

$$k_j^{-1}[\mathbf{M}_t^r]_{js}.$$

Example ([5])



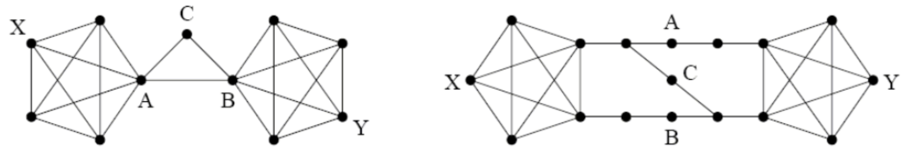
Network 1



Network 2

network		betweenness measure		
		shortest-path	flow	random-walk
Network 1:	vertices A & B	0.636	0.631	0.670
	vertex C	0.200	0.282	0.333
	vertices X & Y	0.200	0.068	0.269
Network 2:	vertices A & B	0.265	0.269	0.321
	vertex C	0.243	0.004	0.267
	vertices X & Y	0.125	0.024	0.194

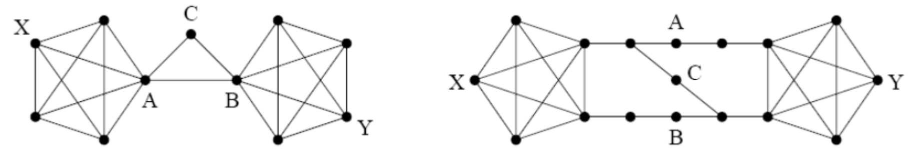
Example ([5])



Network 1 Network 2

network		betweenness measure		
		shortest-path	flow	random-walk
Network 1:	vertices A & B	0.636	0.631	0.670
	vertex C	0.200	0.282	0.333
	vertices X & Y	0.200	0.068	0.269
Network 2:	vertices A & B	0.265	0.269	0.321
	vertex C	0.243	0.004	0.267
	vertices X & Y	0.125	0.024	0.194

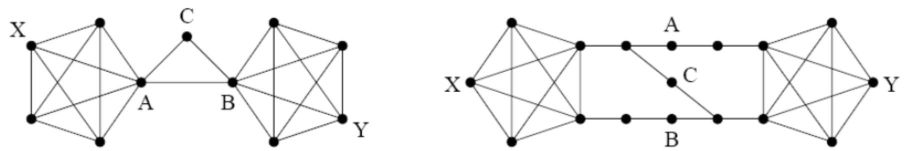
Example ([5])



Network 1 Network 2

network		betweenness measure		
		shortest-path	flow	random-walk
Network 1:	vertices A & B	0.636	0.631	0.670
	vertex C	0.200	0.282	0.333
	vertices X & Y	0.200	0.068	0.269
Network 2:	vertices A & B	0.265	0.269	0.321
	vertex C	0.243	0.004	0.267
	vertices X & Y	0.125	0.024	0.194

Example ([5])



Network 1 Network 2

network		betweenness measure		
		shortest-path	flow	random-walk
Network 1:	vertices A & B	0.636	0.631	0.670
	vertex C	0.200	0.282	0.333
	vertices X & Y	0.200	0.068	0.269
Network 2:	vertices A & B	0.265	0.269	0.321
	vertex C	0.243	0.004	0.267
	vertices X & Y	0.125	0.024	0.194

Feedback-Centrality

- Basic idea: Node is more central the more central its neighbors are.
- Example: Index of Katz:
  - Directed Graph G=(V,E) with edge (a,b) semantics: „a voted for b“
  - Idea: Also count indirect votes; introduce damping function α that gradually lowers contributions from paths with increasing lengths
  - Let $A(k)_{ij}$ denote the number of directed paths from node i to node j of length k;
  - Centrality is then: $c(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{|V|} \alpha(k) A(k)_{ji}$

or in „matrix notation“: $\mathbf{c} = \sum_{k=1}^{\infty} \alpha(k) \mathbf{A}(k)^{\mathsf{T}} (1,1,1,\ldots,1)^{\mathsf{T}} = \boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}} (1,1,1,\ldots,1)^{\mathsf{T}}$

$$(\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{c} = (1,1,1,\ldots,1)^{\mathsf{T}}$$

Feedback-Centrality

- Basic idea: Node is more central the more central its neighbors are.
- Example: Index of Katz:
  - Directed Graph G=(V,E) with edge (a,b) semantics: „a voted for b“
  - Idea: Also count indirect votes; introduce damping function α that gradually lowers contributions from paths with increasing lengths
  - Let $A(k)_{ij}$ denote the number of directed paths from node i to node j of length k;
  - Centrality is then: $c(i)=\sum_{k=1}^{\infty}\sum_{j=1}^{|V|}\alpha(k)A(k)_{ji}$

or in „matrix notation“: $$\mathbf{c}=\sum_{k=1}^{\infty}\alpha(k)\mathbf{A}(k)^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}=\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}$$
$$(\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{c}=(1,1,1,\ldots,1)^{\mathsf{T}}$$

Feedback-Centrality

- Basic idea: Node is more central the more central its neighbors are.
- Example: Index of Katz:
  - Directed Graph G=(V,E) with edge (a,b) semantics: „a voted for b“
  - Idea: Also count indirect votes; introduce damping function α that gradually lowers contributions from paths with increasing lengths
  - Let $A(k)_{ij}$ denote the number of directed paths from node i to node j of length k;
  - Centrality is then: $c(i)=\sum_{k=1}^{\infty}\sum_{j=1}^{|V|}\alpha(k)A(k)_{ji}$

or in „matrix notation“: $$\mathbf{c}=\sum_{k=1}^{\infty}\alpha(k)\mathbf{A}(k)^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}=\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}$$
$$(\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{c}=(1,1,1,\ldots,1)^{\mathsf{T}}$$

Feedback-Centrality

- Basic idea: Node is more central the more central its neighbors are.
- Example: Index of Katz:
  - Directed Graph G=(V,E) with edge (a,b) semantics: „a voted for b“
  - Idea: Also count indirect votes; introduce damping function α that gradually lowers contributions from paths with increasing lengths
  - Let $A(k)_{ij}$ denote the number of directed paths from node i to node j of length k;
  - Centrality is then: $c(i)=\sum_{k=1}^{\infty}\sum_{j=1}^{|V|}\alpha(k)A(k)_{ji}$

or in „matrix notation“: $$\mathbf{c}=\sum_{k=1}^{\infty}\alpha(k)\mathbf{A}(k)^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}=\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}$$
$$(\boldsymbol{\alpha}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{c}=(1,1,1,\ldots,1)^{\mathsf{T}}$$

Feedback-Centrality

A is the usual adjacency matrix of G

- If $\alpha(k) = \alpha^k$ and by observing that $A(k)_{ij} = (A^k)_{ij}$ and assuming convergence of the geometric series, the equations become

$$\mathbf{c}=\sum_{k=1}^{\infty}\alpha^{k}(\mathbf{A}^{k})^{\mathsf{T}}(1,1,1,\ldots,1)^{\mathsf{T}}=\mathbf{I}(\mathbf{I}-\alpha\mathbf{A}^{\mathsf{T}})^{-1}(1,1,1,\ldots,1)^{\mathsf{T}}$$

where **I** is the identity matrix.
Thus we have:

$$(\mathbf{I}-\alpha\,\mathbf{A}^{\mathsf{T}})\mathbf{c}=(1,1,1,\ldots,1)^{\mathsf{T}}$$

which shows that centrality values depend on each other.

If the largest eigenvalue of A is less than 1/α then the series converges

Feedback-Centrality

A is the usual adjacency matrix of G

- If $\alpha(k) = \alpha^k$ and by observing that $A(k)_{ij} = (A^k)_{ij}$ and assuming convergence of the geometric series, the equations become

$$\mathbf{c} = \sum_{k=1}^{\infty} \alpha^k (\mathbf{A}^k)^{\mathbf{T}} (1,1,1,\ldots,1)^{\mathbf{T}} = \mathbf{I}(\mathbf{I} - \alpha \mathbf{A}^{\mathbf{T}})^{-1} (1,1,1,\ldots,1)^{\mathbf{T}}$$

where **I** is the identity matrix.
Thus we have:

$$(\mathbf{I} - \alpha\, \mathbf{A}^{\mathbf{T}})\mathbf{c} = (1,1,1,\ldots,1)^{\mathbf{T}}$$

which shows that centrality values depend on each other.

If the largest eigenvalue of A is less than 1/α then the series converges

Feedback-Centrality

A is the usual adjacency matrix of G

- If $\alpha(k) = \alpha^k$ and by observing that $A(k)_{ij} = (A^k)_{ij}$ and assuming convergence of the geometric series, the equations become

$$\mathbf{c} = \sum_{k=1}^{\infty} \alpha^k (\mathbf{A}^k)^{\mathbf{T}} (1,1,1,\ldots,1)^{\mathbf{T}} = \mathbf{I}(\mathbf{I} - \alpha \mathbf{A}^{\mathbf{T}})^{-1} (1,1,1,\ldots,1)^{\mathbf{T}}$$

where **I** is the identity matrix.
Thus we have:

$$(\mathbf{I} - \alpha\, \mathbf{A}^{\mathbf{T}})\mathbf{c} = (1,1,1,\ldots,1)^{\mathbf{T}}$$

which shows that centrality values depend on each other.

If the largest eigenvalue of A is less than 1/α then the series converges

Feedback-Centrality

A is the usual adjacency matrix of G

- If $\alpha(k) = \alpha^k$ and by observing that $A(k)_{ij} = (A^k)_{ij}$ and assuming convergence of the geometric series, the equations become

$$\mathbf{c} = \sum_{k=1}^{\infty} \alpha^k (\mathbf{A}^k)^{\mathbf{T}} (1,1,1,\ldots,1)^{\mathbf{T}} = \mathbf{I}(\mathbf{I} - \alpha \mathbf{A}^{\mathbf{T}})^{-1} (1,1,1,\ldots,1)^{\mathbf{T}}$$

where **I** is the identity matrix.
Thus we have:

$$(\mathbf{I} - \alpha\, \mathbf{A}^{\mathbf{T}})\mathbf{c} = (1,1,1,\ldots,1)^{\mathbf{T}}$$

which shows that centrality values depend on each other.

If the largest eigenvalue of A is less than 1/α then the series converges

Feedback-Centrality

A is the usual adjacency matrix of G

- If $\alpha(k) = \alpha^k$ and by observing that $A(k)_{ij} = (A^k)_{ij}$ and assuming convergence of the geometric series, the equations become

$$\mathbf{c} = \sum_{k=1}^{\infty} \alpha^k (\mathbf{A}^k)^{\mathbf{T}} (1,1,1,\ldots,1)^{\mathbf{T}} = \mathbf{I}(\mathbf{I} - \alpha \mathbf{A}^{\mathbf{T}})^{-1} (1,1,1,\ldots,1)^{\mathbf{T}}$$

where **I** is the identity matrix.
Thus we have:

$$(\mathbf{I} - \alpha\, \mathbf{A}^{\mathbf{T}})\mathbf{c} = (1,1,1,\ldots,1)^{\mathbf{T}}$$

which shows that centrality values depend on each other.

If the largest eigenvalue of A is less than 1/α then the series converges

Feedback-Centrality

Further example: Hubbell index

- weighted, directed graph G=(V,E); weights formalized in adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with edge weight connecting these nodes to v).
- centrality vector s of the nodes is thus an eigenvector of **W: s=Ws**
- In order to make this equation solvable, introduce a „centrality input“ or „external information“ E(v) for every node v: **s**=**E**+**Ws**
- → **s=(I-W)$^{-1}$E**
- **I-W** is invertible if $\sum_{k=1}^{\infty} W^k$ converges ←→the largest eigenvalue of W is less than one (see[1]).

Feedback-Centrality

Further example: Hubbell index

- weighted, directed graph G=(V,E); weights formalized in adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with edge weight connecting these nodes to v).
- centrality vector s of the nodes is thus an eigenvector of **W: s=Ws**
- In order to make this equation solvable, introduce a „centrality input“ or „external information“ E(v) for every node v: **s**=**E**+**Ws**
- → **s=(I-W)$^{-1}$E**
- **I-W** is invertible if $\sum_{k=1}^{\infty} W^k$ converges ←→the largest eigenvalue of W is less than one (see[1]).

Feedback-Centrality

Further example: Hubbell index

- weighted, directed graph G=(V,E); weights formalized in adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with edge weight connecting these nodes to v).
- centrality vector s of the nodes is thus an eigenvector of **W: s=Ws**
- In order to make this equation solvable, introduce a „centrality input“ or „external information“ E(v) for every node v: **s**=**E**+**Ws**
- → **s=(I-W)$^{-1}$E**
- **I-W** is invertible if $\sum_{k=1}^{\infty} W^k$ converges ←→the largest eigenvalue of W is less than one (see[1]).

Feedback-Centrality

Further example: Hubbell index

- weighted, directed graph G=(V,E); weights formalized in adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with edge weight connecting these nodes to v).
- centrality vector s of the nodes is thus an eigenvector of **W: s=Ws**
- In order to make this equation solvable, introduce a „centrality input“ or „external information“ E(v) for every node v: **s**=**E**+**Ws**
- → **s=(I-W)$^{-1}$E**
- **I-W** is invertible if $\sum_{k=1}^{\infty} W^k$ converges ←→the largest eigenvalue of W is less than one (see[1]).

Feedback-Centrality

Further example: Hubbell index

- weighted, directed graph G=(V,E); weights formalized in adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with edge weight connecting these nodes to v).
- centrality vector s of the nodes is thus an eigenvector of **W: s=Ws**
- In order to make this equation solvable, introduce a „centrality input“ or „external information“ E(v) for every node v: **s=E+Ws**
- → **s=(I-W)**$^{-1}$**E**
- **I-W** is invertible if $\sum_{k=1}^{\infty} W^k$ converges ←→the largest eigenvalue of W is less than one (see[1]).

Feedback-Centrality

- Further example: Random surfer on Web-pages
- Directed unweighted graph G=(V,E)
- Define Markov transition matrix as

$$t_{ij} = \begin{cases} \dfrac{1}{\deg^+(i)} & if\ (i,j) \in E \\ 0 & if\ (i,j) \notin E \\ \dfrac{1}{|V|} & if\ \deg^+(i) = 0 \end{cases}$$

(choose one outgoing link randomly, probability inverse propotional to out degree of current node; if node is a sink (no outgoing links) choose a random page)

Feedback-Centrality

- Further example: Random surfer on Web-pages
- Directed unweighted graph G=(V,E)
- Define Markov transition matrix as

$$t_{ij} = \begin{cases} \dfrac{1}{\deg^+(i)} & if\ (i,j) \in E \\ 0 & if\ (i,j) \notin E \\ \dfrac{1}{|V|} & if\ \deg^+(i) = 0 \end{cases}$$

(choose one outgoing link randomly, probability inverse propotional to out degree of current node; if node is a sink (no outgoing links) choose a random page)

Feedback-Centrality

- Question: is there a unique stationary distribution π? (→ in essence is the chain irreducible and positively recurrent?)
- → make it irreducible: T=αT+(1- α)E where E is the matrix with all entries equal to 1/n (completely stochastic choosing).
- social analog: „assigning leadership“, „seeking friends“; „expert seeking“ etc.
- Stationary distributions ←→ degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v \in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i \in V} \pi_i t_{ij} = \frac{\sum_{i \in V} \deg(i) t_{ij}}{\sum_{v \in V} \deg(v)} = \frac{\sum_{i \in V} A_{ij}}{\sum_{v \in V} \deg(v)} = \frac{\deg(j)}{\sum_{v \in V} \deg(v)} = \pi_j$$

Feedback-Centrality

- Question: is there a unique stationary distribution π? (→ in essence is the chain irreducible and positively recurrent?)
- → make it irreducible: T=αT+(1- α)E where E is the matrix with all entries equal to 1/n (completely stochastic choosing).
- social analog: „assigning leadership“, „seeking friends“; „expert seeking“ etc.
- Stationary distributions ←→ degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v\in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i\in V} \pi_i t_{ij} = \frac{\sum_{i\in V} \deg(i) t_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\sum_{i\in V} A_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\deg(j)}{\sum_{v\in V} \deg(v)} = \pi_j$$

Feedback-Centrality

- Question: is there a unique stationary distribution π? (→ in essence is the chain irreducible and positively recurrent?)
- → make it irreducible: T=αT+(1- α)E where E is the matrix with all entries equal to 1/n (completely stochastic choosing).
- social analog: „assigning leadership“, „seeking friends“; „expert seeking“ etc.
- Stationary distributions ←→ degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v\in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i\in V} \pi_i t_{ij} = \frac{\sum_{i\in V} \deg(i) t_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\sum_{i\in V} A_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\deg(j)}{\sum_{v\in V} \deg(v)} = \pi_j$$

Feedback-Centrality

- Question: is there a unique stationary distribution π? (→ in essence is the chain irreducible and positively recurrent?)
- → make it irreducible: T=αT+(1- α)E where E is the matrix with all entries equal to 1/n (completely stochastic choosing).
- social analog: „assigning leadership“, „seeking friends“; „expert seeking“ etc.
- Stationary distributions ←→ degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v\in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i\in V} \pi_i t_{ij} = \frac{\sum_{i\in V} \deg(i) t_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\sum_{i\in V} A_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\deg(j)}{\sum_{v\in V} \deg(v)} = \pi_j$$

Feedback-Centrality

- Question: is there a unique stationary distribution π? (→ in essence is the chain irreducible and positively recurrent?)
- → make it irreducible: T=αT+(1- α)E where E is the matrix with all entries equal to 1/n (completely stochastic choosing).
- social analog: „assigning leadership“, „seeking friends“; „expert seeking“ etc.
- Stationary distributions ←→ degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v\in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i\in V} \pi_i t_{ij} = \frac{\sum_{i\in V} \deg(i) t_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\sum_{i\in V} A_{ij}}{\sum_{v\in V} \deg(v)} = \frac{\deg(j)}{\sum_{v\in V} \deg(v)} = \pi_j$$

Feedback-Centrality: Page Rank

- Famous ingredient of Google
- Centrality of a web-page depends on the centralities of the pages linking to it:

$$c(p) = d \sum_{q \in \{\text{"In-neighbors of } p\text{"}\} = \Gamma^-(p)} \frac{c(q)}{\deg^+(q)} + (1-d)$$

where d is a damping factor; $\deg^+(q)$ is the out degree of q.

- Matrix Notation:

$$\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$$

where transition matrix $P_{ij} = 1/\deg^+(j)$ if $(j,i) \in E$ and $P_{ij} = 0$ otherwise

Feedback-Centrality: Page Rank

- Famous ingredient of Google
- Centrality of a web-page depends on the centralities of the pages linking to it:

$$c(p) = d \sum_{q \in \{\text{"In-neighbors of } p\text{"}\} = \Gamma^-(p)} \frac{c(q)}{\deg^+(q)} + (1-d)$$

where d is a damping factor; $\deg^+(q)$ is the out degree of q.

- Matrix Notation:

$$\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$$

where transition matrix $P_{ij} = 1/\deg^+(j)$ if $(j,i) \in E$ and $P_{ij} = 0$ otherwise

Feedback-Centrality: Page Rank

- Solving the equation $\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$:
- If $0 \le d < 1$ the equation has a unique solution

$$\mathbf{c} = (1-d)(\mathbf{I} - d\,\mathbf{P})^{-1}(1,1,\ldots,1)^{\mathsf{T}}$$

- How do we compute the solution avoiding matrix inversion? → Jacobi power iteration:

$$c_i^{(k+1)} = d \sum_j P_{ij} c_j^{(k)} + (1-d)$$

or improved variant (Gauss-Seidel iteration): (see [3])

$$c_i^{(k+1)} = d\,\Big(\sum_{j<i} P_{ij} c_j^{(k+1)} + \sum_{j \ge i} P_{ij} c_j^{(k)}\Big) + (1-d)$$

Feedback-Centrality: Page Rank

- Solving the equation $\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$:
- If $0 \le d < 1$ the equation has a unique solution

$$\mathbf{c} = (1-d)(\mathbf{I} - d\,\mathbf{P})^{-1}(1,1,\ldots,1)^{\mathsf{T}}$$

- How do we compute the solution avoiding matrix inversion? → Jacobi power iteration:

$$c_i^{(k+1)} = d \sum_j P_{ij} c_j^{(k)} + (1-d)$$

or improved variant (Gauss-Seidel iteration): (see [3])

$$c_i^{(k+1)} = d\,\Big(\sum_{j<i} P_{ij} c_j^{(k+1)} + \sum_{j \ge i} P_{ij} c_j^{(k)}\Big) + (1-d)$$

Feedback-Centrality: Page Rank

- Solving the equation $\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$:
- If 0 ≤ d <1 the equation has a unique solution

$$\mathbf{c} = (1-d)(\mathbf{I} - d\,\mathbf{P})^{-1}(1,1,\ldots,1)^{\mathsf{T}}$$

- How do we compute the solution avoiding matrix inversion? → Jacobi power iteration:

$$c_i^{(k+1)} = d \sum_j P_{ij} c_j^{(k)} + (1-d)$$

or improved variant (Gauss-Seidel iteration): (see [3])

$$c_i^{(k+1)} = d\,\Big(\sum_{j<i} P_{ij} c_j^{(k+1)} + \sum_{j\geq i} P_{ij} c_j^{(k)}\Big) + (1-d)$$

Feedback-Centrality: Page Rank

- Solving the equation $\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$:
- If 0 ≤ d <1 the equation has a unique solution

$$\mathbf{c} = (1-d)(\mathbf{I} - d\,\mathbf{P})^{-1}(1,1,\ldots,1)^{\mathsf{T}}$$

- How do we compute the solution avoiding matrix inversion? → Jacobi power iteration:

$$c_i^{(k+1)} = d \sum_j P_{ij} c_j^{(k)} + (1-d)$$

or improved variant (Gauss-Seidel iteration): (see [3])

$$c_i^{(k+1)} = d\,\Big(\sum_{j<i} P_{ij} c_j^{(k+1)} + \sum_{j\geq i} P_{ij} c_j^{(k)}\Big) + (1-d)$$

Feedback-Centrality: Page Rank

- Solving the equation $\mathbf{c} = d\,\mathbf{P}\mathbf{c} + (1-d)(1,1,\ldots,1)^{\mathsf{T}}$:
- If 0 ≤ d <1 the equation has a unique solution

$$\mathbf{c} = (1-d)(\mathbf{I} - d\,\mathbf{P})^{-1}(1,1,\ldots,1)^{\mathsf{T}}$$

- How do we compute the solution avoiding matrix inversion? → Jacobi power iteration:

$$c_i^{(k+1)} = d \sum_j P_{ij} c_j^{(k)} + (1-d)$$

or improved variant (Gauss-Seidel iteration): (see [3])

$$c_i^{(k+1)} = d\,\Big(\sum_{j<i} P_{ij} c_j^{(k+1)} + \sum_{j\geq i} P_{ij} c_j^{(k)}\Big) + (1-d)$$

N-Cliques, N-Clubs, N-Clans

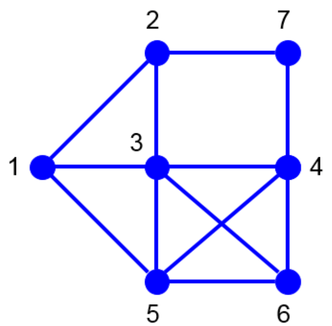
- Cliques are very “strict” → Alternative candidates for groups: Distance based structures:
  - U is N-clique iff ∀u,v ∈ U : $\mathrm{dist}_G(u,v) \leq N$ (non-local def.!)
  - U is N-club iff diam(G([U])) ≤ N
  - U is N-clan iff U is maximal N-clique and diam(G([U])) ≤ N
- Criticisms:
  - Since dist is evaluated w.r.t. to G and not G([U]) (thus N-cliques are not local structures), N-cliques need not even be connected and can have a diameter diam(G([U])) > N

- Cliques are very "strict" → Alternative candidates for groups: Distance based structures:
  - U is N-clique iff $\forall u,v \in U$: $dist_G(u,v) \le N$ (non-local def.!)
  - U is N-club iff $diam(G([U])) \le N$
  - U is N-clan iff U is maximal N-clique and $diam(G([U])) \le N$
- Criticisms:
  - Since dist is evaluated w.r.t. to G and not G([U]) (thus N-cliques are not local structures), N-cliques need not even be connected and can have a diameter $diam(G([U])) > N$

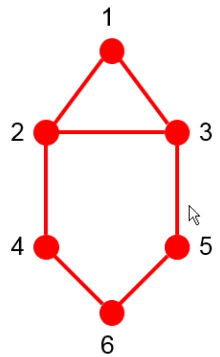
- U is N-clique iff $\forall u,v \in V$: $dist_G(u,v) \le N$
- U is N-club iff $diam(G([U])) \le N$
- U is N-clan iff U is maximal N-clique and $diam(G([U])) \le N$

- → N-clan: restrict dist-condition to paths of nodes within the structure: easy to find (just drop all n-cliques with diameter greater than N)
- → N-club: regard all induced graphs with diameter less than N: harder to find
- It can be shown / seen from the def.:
  - -- all N-clans are N-cliques;
  - -- all N-clubs are contained within N-cliques;
  - -- all N-clans are n-clubs
  - -- there are N-clubs that are not N-clans

- U is N-clique iff $\forall u,v \in V$: $dist_G(u,v) \le N$
- U is N-club iff $diam(G([U])) \le N$
- U is N-clan iff U is maximal N-clique and $diam(G([U])) \le N$



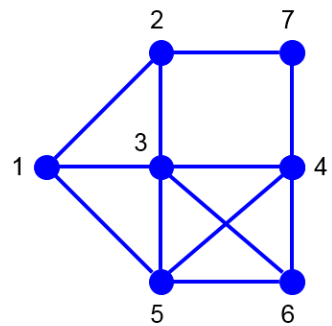
cliques: {1, 2, 3}, {1, 3, 5}, {3, 4, 5, 6}



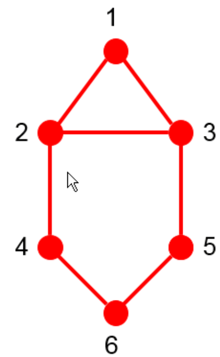
2-cliques: {1, 2, 3, 4, 5}, {2, 3, 4, 5, 6}
2-clubs: {1, 2, 3, 4}, {1, 2, 3, 5}, {2, 3, 4, 5, 6}
2-clan: {2, 3, 4, 5, 6}

N-Cliques, N-Clubs, N-Clans

- U is N-clique iff $\forall u,v \in V$: $dist_G(u,v) \le N$
- U is N-club iff $diam(G([U])) \le N$
- U is N-clan iff U is maximal N-clique and $diam(G([U])) \le N$



cliques: {1, 2, 3}, {1, 3, 5}, {3, 4, 5, 6}



2-cliques: {1, 2, 3, 4, 5}, {2, 3, 4, 5, 6}
2-clubs: {1, 2, 3, 4}, {1, 2, 3, 5}, {2, 3, 4, 5, 6}
2-clan: {2, 3, 4, 5, 6}

N-Cliques, N-Clubs, N-Clans

- U is N-clique iff $\forall u,v \in U$: $dist_G(u,v) \leq N$
- U is N-club iff $diam(G([U])) \leq N$
- U is N-clan iff U is maximal N-clique and $diam(G([U])) \leq N$

- Further criticism:
  - Small distances are characteristic even for large social networks (cmp. 6 degrees) → N-cliques, N-clubs and N-clans may not be socially meaningful as groups but may be interesting for modeling social influence/neighbourhood spheres (e.g. regarding information flows (compare [13], p. 263))
  - These constructs are not generally closed under exclusion and are not nested (socially meaningful characteristics that cliques possess)