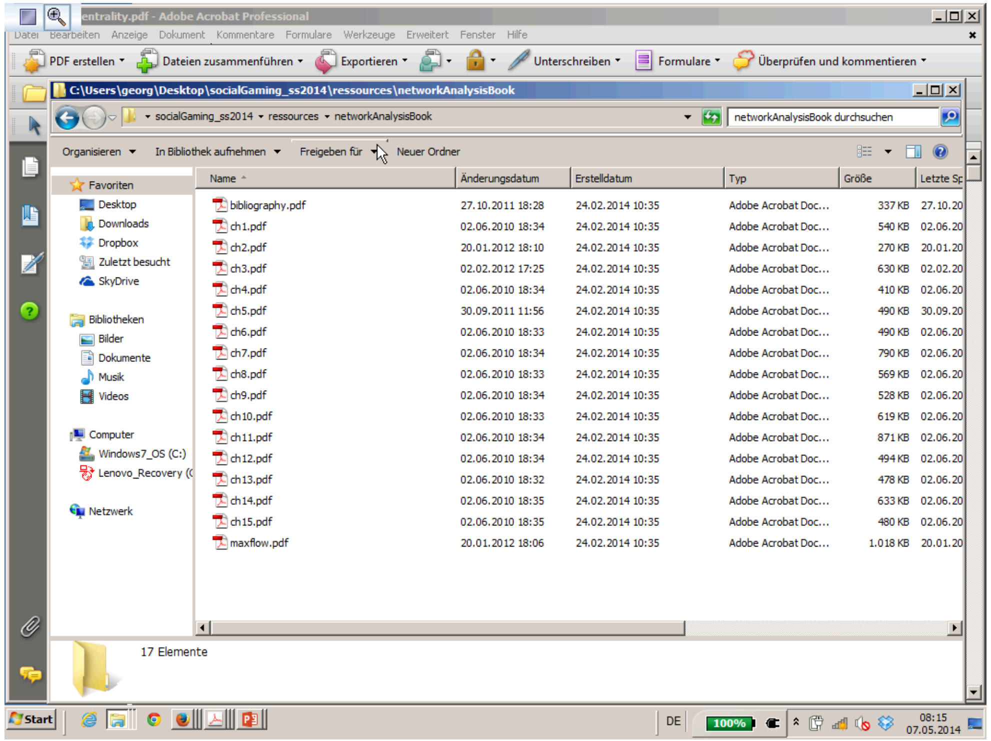
Script **generated by TTT**

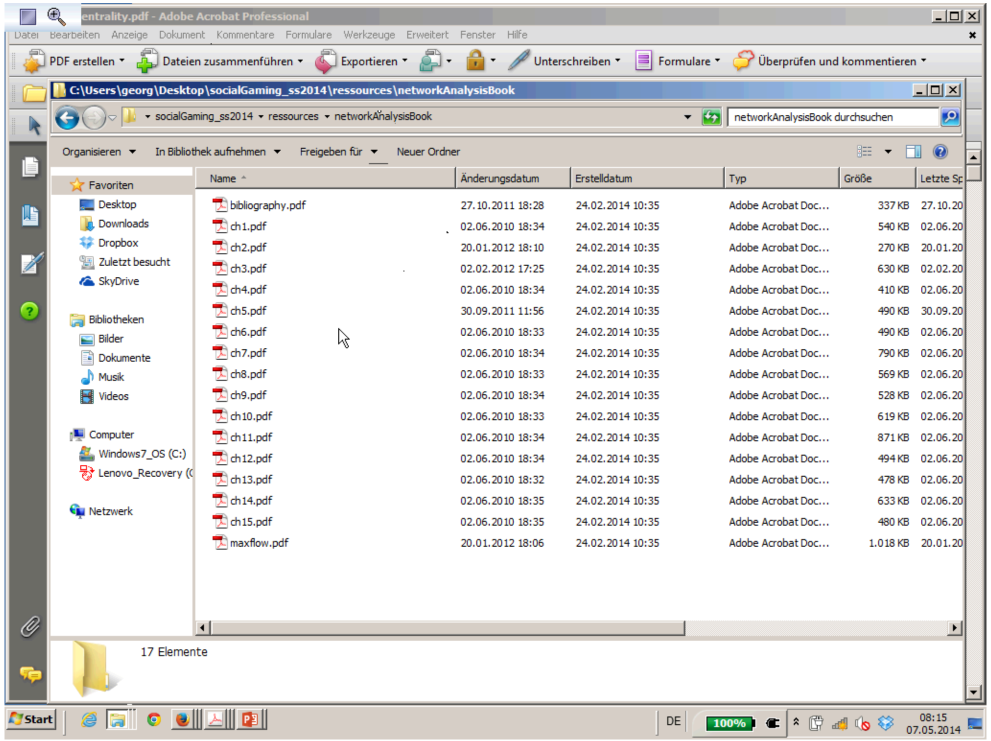
Title: groh: profile1 (07.05.2014)

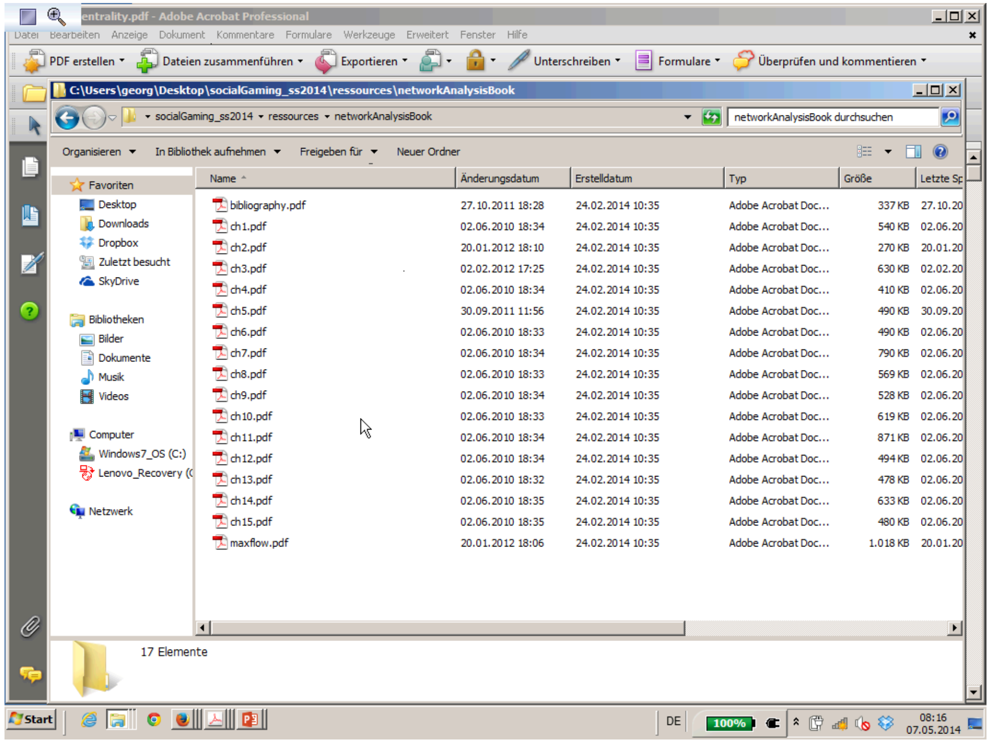
Date: Wed May 07 08:15:29 CEST 2014

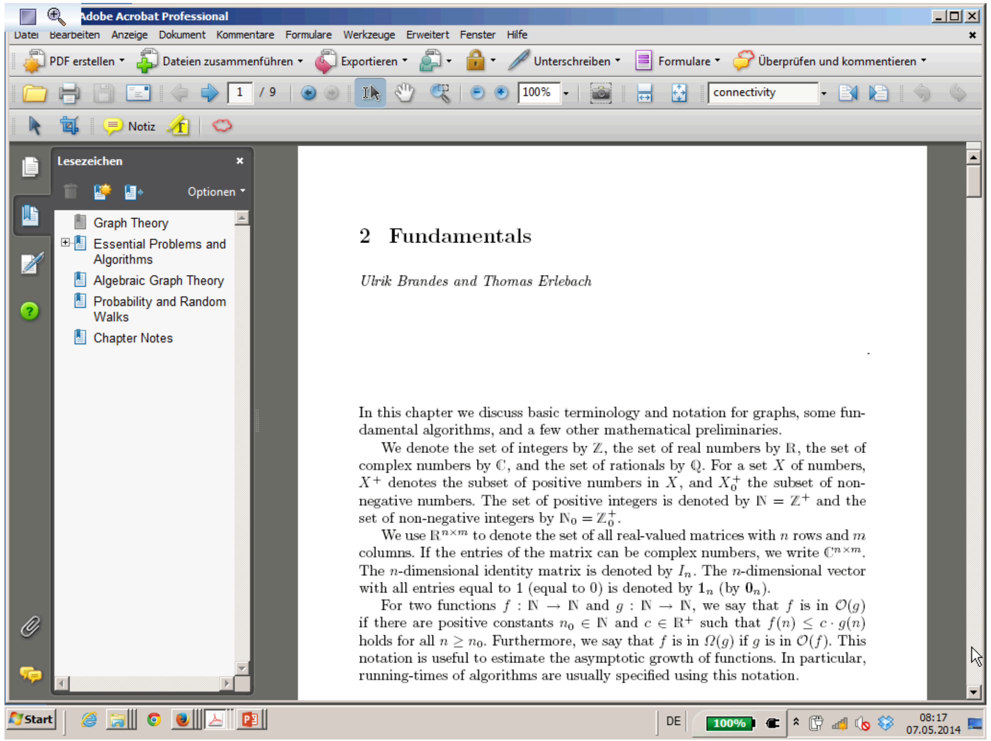
Duration: 86:26 min

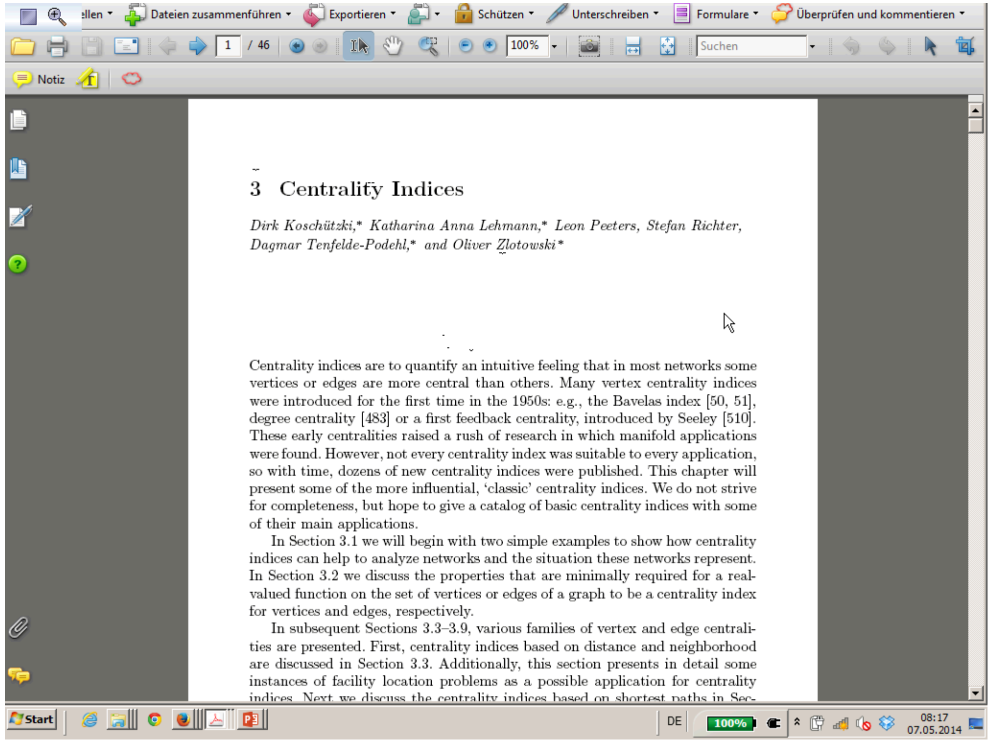
Pages: 77

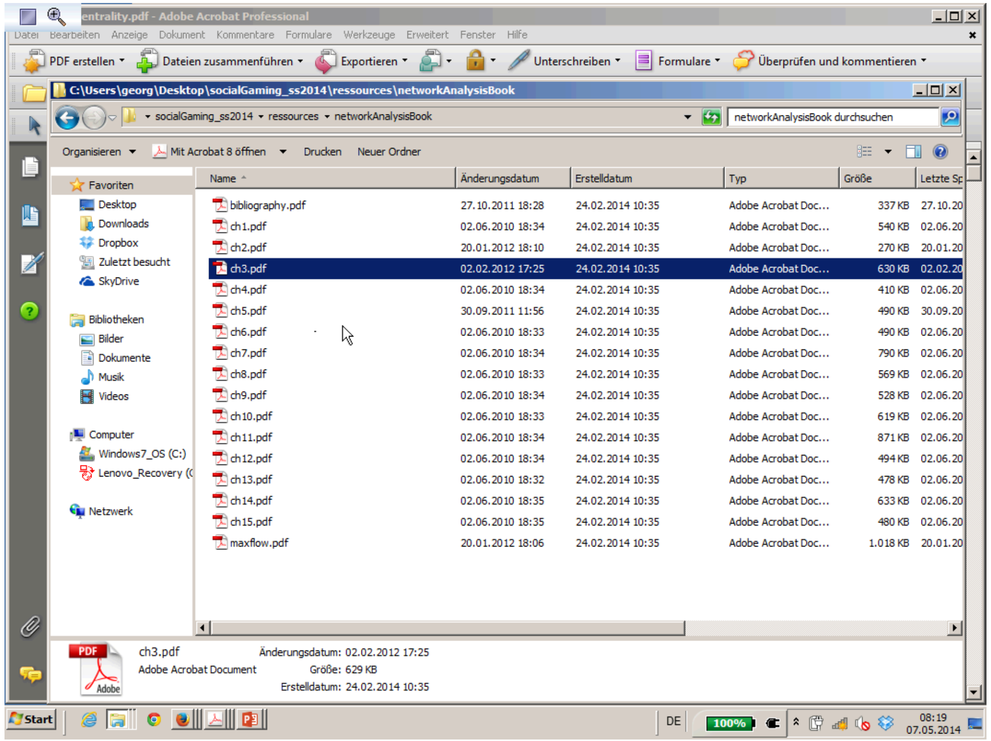








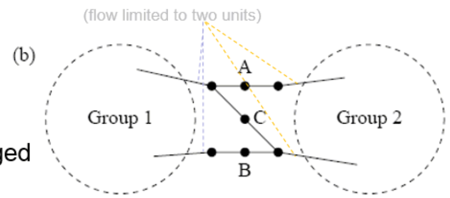




Critique on Betweeness Based Centralities

- 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.

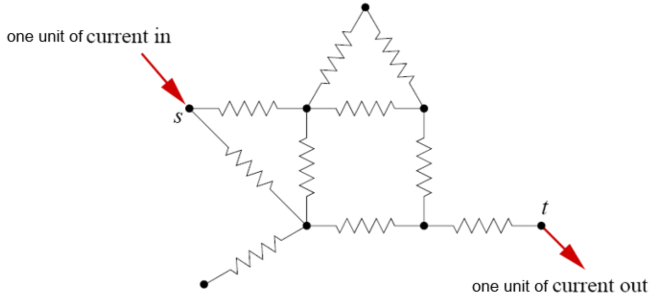
Source: [5]

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

- random walk based centrality rwb: idea:
  rwb(i) := number of times that a random walk starting at s and ending at t passes through i along the way, averaged over all s and t
- rwb ↔ spb: opposite ends:
  - rwb: info has no idea where its going
  - spb: info knows exactly where its going
- compute for all i rwb(i): $O((m+n)n^2)$ worst case time complexity (comparable to spb)

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

- flow of electric current in a resistor network;
  V_i = voltage (potential) at vertex i
- ↔ Current Flow betweenness cfb centrality : cfb(i) := amount of current that flows through i in this setup, averaged over all s and t.



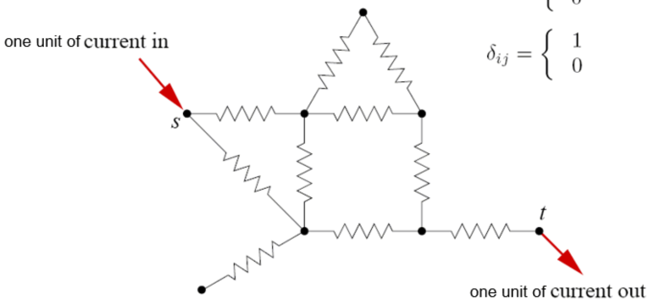
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



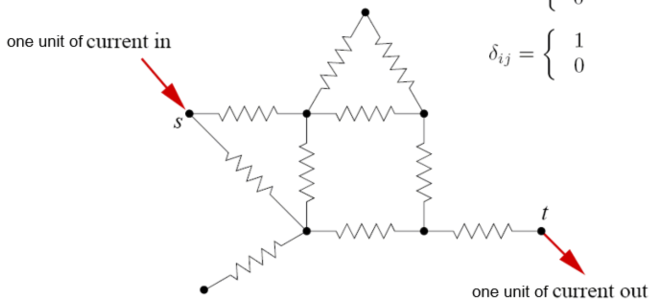
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



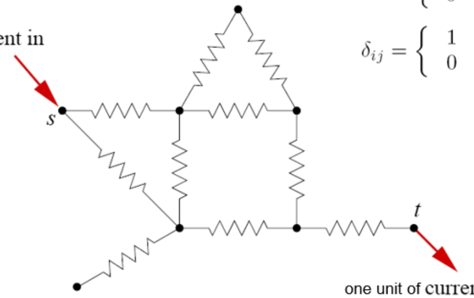
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



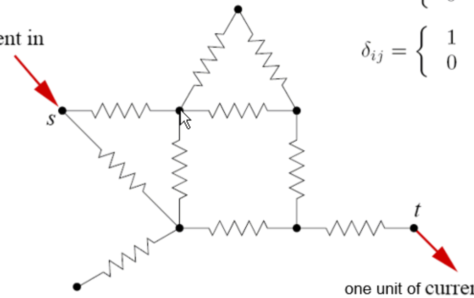
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



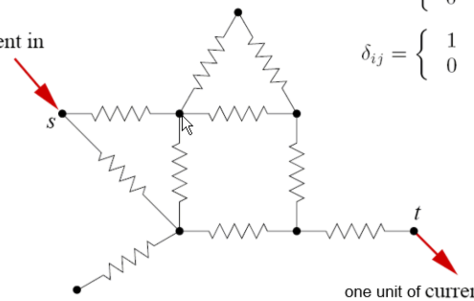
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



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

$\sum_j A_{ij} = k_i$, the degree of vertex i

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \longrightarrow \underbrace{(\mathbf{D} - \mathbf{A})}_{\text{“Graph Laplacian”}} \cdot \mathbf{V} = \mathbf{s}$$

$\mathbf{D}$ is the diagonal matrix with elements $D_{ii} = k_i$

$$\text{source vector } \mathbf{s} \quad s_i = \begin{cases} +1 & \text{for } i = s, \\ -1 & \text{for } i = t, \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}$$

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

$\sum_j A_{ij} = k_i$, the degree of vertex i

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \longrightarrow \underbrace{(\mathbf{D} - \mathbf{A})}_{\text{"Graph Laplacian"}} \cdot \mathbf{V} = \mathbf{s}$$

$\mathbf{D}$ is the diagonal matrix with elements $D_{ii} = k_i$

source vector $\mathbf{s}$ $\quad s_i = \begin{cases} +1 & \text{for } i = s, \\ -1 & \text{for } i = t, \\ 0 & \text{otherwise.} \end{cases}$

$$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}$$

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

$\sum_j A_{ij} = k_i$, the degree of vertex i

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \longrightarrow \underbrace{(\mathbf{D} - \mathbf{A})}_{\text{"Graph Laplacian"}} \cdot \mathbf{V} = \mathbf{s}$$

$\mathbf{D}$ is the diagonal matrix with elements $D_{ii} = k_i$

source vector $\mathbf{s}$ $\quad s_i = \begin{cases} +1 & \text{for } i = s, \\ -1 & \text{for } i = t, \\ 0 & \text{otherwise.} \end{cases}$

$$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}$$

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

$\sum_j A_{ij} = k_i$, the degree of vertex i

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \longrightarrow \underbrace{(\mathbf{D} - \mathbf{A})}_{\text{"Graph Laplacian"}} \cdot \mathbf{V} = \mathbf{s}$$

$\mathbf{D}$ is the diagonal matrix with elements $D_{ii} = k_i$

source vector $\mathbf{s}$ $\quad s_i = \begin{cases} +1 & \text{for } i = s, \\ -1 & \text{for } i = t, \\ 0 & \text{otherwise.} \end{cases}$

$$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}$$

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

$$\underbrace{(\mathbf{D} - \mathbf{A})} \cdot \mathbf{V} = \mathbf{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}$ (matrix inversion: $O(n^3)$)

let us now add a vth row and column back into $(\mathbf{D}_v - \mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$\longrightarrow V_i^{(st)} = T_{is} - T_{it}$

$\longrightarrow$ current flow at node i: $I_i^{(st)} = \frac{1}{2} \sum_j A_{ij} |V_i^{(st)} - V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n^3))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\quad V_i^{(st)}=T_{is}-T_{it}$$

$\longrightarrow$ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n^3))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\quad V_i^{(st)}=T_{is}-T_{it}$$

$\longrightarrow$ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n^3))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\quad V_i^{(st)}=T_{is}-T_{it}$$

$\longrightarrow$ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n^3))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\quad V_i^{(st)}=T_{is}-T_{it}$$

$\longrightarrow$ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n³))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\ V_i^{(st)}=T_{is}-T_{it}$$

⟶ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n³))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\ V_i^{(st)}=T_{is}-T_{it}$$

⟶ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

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

$$\underbrace{(\mathbf{D}-\mathbf{A})}\cdot\mathbf{V}=\mathrm{s}$$

Laplacian is not invertible, det = 0, because system of eq. is overdetermined → set one V_v=0 and measure voltages relative to v. → remove the v-th row and column (since V_v=0) → now invertible

$$\mathbf{V}=(\mathbf{D}_v-\mathbf{A}_v)^{-1}\cdot\mathrm{s}$$ (matrix inversion: O(n³))

let us now add a vth row and column back into $(\mathbf{D}_v-\mathbf{A}_v)^{-1}$ with values all equal to zero.
The resulting matrix we will denote $\mathbf{T}$.

$$\longrightarrow\ V_i^{(st)}=T_{is}-T_{it}$$

⟶ current flow at node i: $I_i^{(st)}=\frac{1}{2}\sum_j A_{ij}|V_i^{(st)}-V_j^{(st)}|$

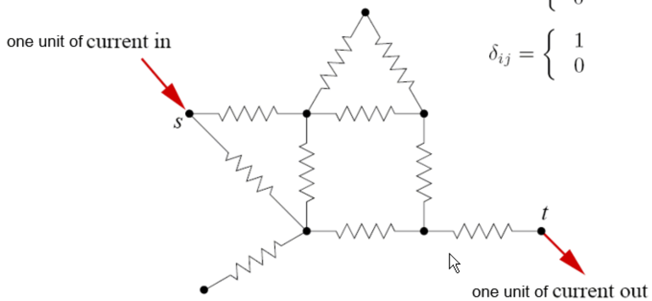
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i-V_j)=\delta_{is}-\delta_{it},$$

$$A_{ij}=\begin{cases}1 & \text{if there is an edge between } i \text{ and } j,\\ 0 & \text{otherwise,}\end{cases}$$

$$\delta_{ij}=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{otherwise.}\end{cases}$$



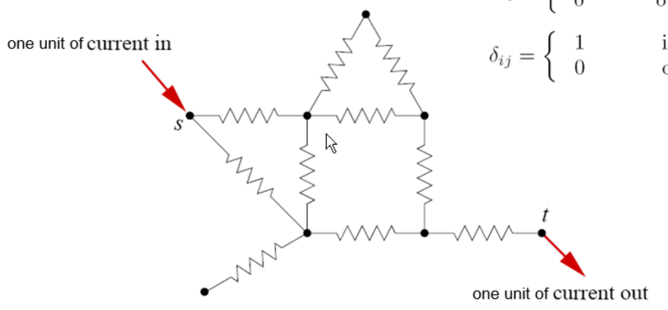
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- Kirchhoffs point law (current conservation): total current flow in / out of node is zero:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it},$$

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$



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])

- 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])

- 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 node j, probability that in next step at node i is:

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

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-1}$$

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 node j, probability that in next step at node i is:

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

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

$$D_{ii} = k_i$$

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

- 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 node j, probability that in next step at node i is:

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

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

$$D_{ii} = k_i$$

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

- 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 node j, probability that in next step at node i is:

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

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-1}$$ 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 →

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

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

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

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 →

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

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

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

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 →

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

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

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

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 →

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

- 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;

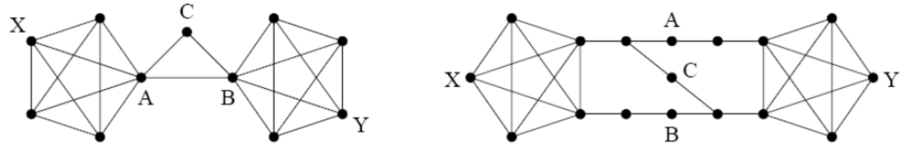
denote by $\mathbf{M}_t = \mathbf{A}_t \cdot \mathbf{D}_t^{-1}$ the matrix with these elements removed, and similarly for A_t and D_t.

- 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

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

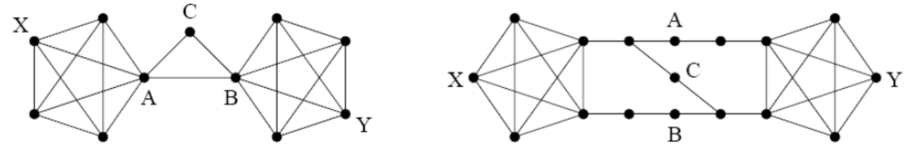
Example ([5])



Network 1 Network 2

network		betweenness measure: shortest-path	max-flow	random walk / current-flow
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

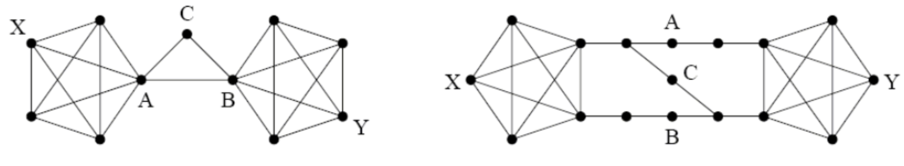
Example ([5])



Network 1 Network 2

network		betweenness measure: shortest-path	max-flow	random walk / current-flow
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	max-flow	random walk / current-flow
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: Hubbell index

- weighted, directed graph G=(V,E): weighted adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with corresp. edge weight). → 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

Basic idea: Node is more central the more central its neighbors are.

example: Hubbell index

- weighted, directed graph G=(V,E): weighted adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with corresp. edge weight). → 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

Basic idea: Node is more central the more central its neighbors are.

example: Hubbell index

- weighted, directed graph G=(V,E): weighted adjacency matrix **W**
- centralilty s(v) of node v is proportional to sum of centralities s(w) of adjacent nodes w (multiplied with corresp. edge weight). → 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]).

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 →

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

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

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

Feedback-Centrality

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

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

(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

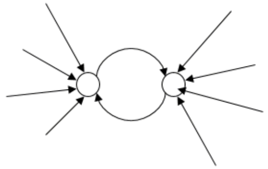
- Further example: Random surfer on Web-pages
- Directed unweighted graph G=(V,E)
- Define Markov transition matrix as

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

(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

In order to avoid getting stuck in “sink circles”, we can add a small probability here of choosing randomly. After that we have to renormalize to keep the matrix T stochastic.



- Fu
- Di
- De

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

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

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

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

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

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 choice).
- 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:

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

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:

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

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:

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

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

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

- 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])

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

Feedback-Centrality: Page Rank

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

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

- 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])

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

Feedback-Centrality: Page Rank

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

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

- 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])

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

Feedback-Centrality: Page Rank

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

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

- 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])

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

Recommended Reading

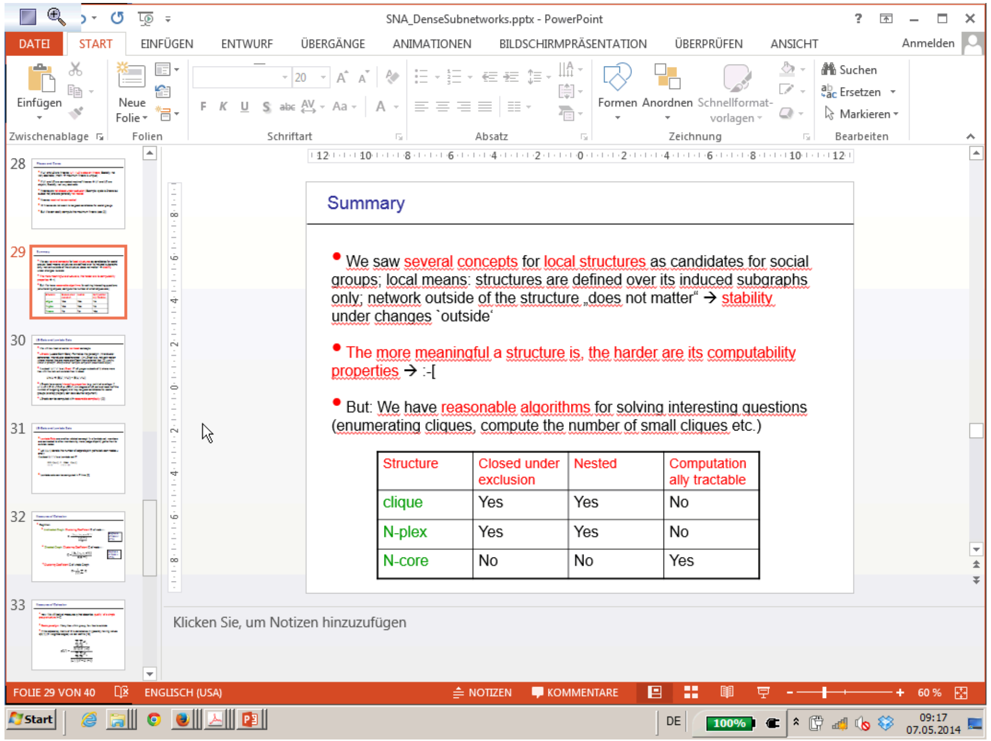
- minimal approach:
  - study the slides and mentally review the introduced concepts, definitions and connections
- standard approach:
  - minimal approach + read the corresponding parts of [1] and [5]
- interested students:
  - standard approach + read [4]

students with problems w.r.t. graph theory: read [2]

Recommended Reading

- minimal approach:
  - study the slides and mentally review the introduced concepts, definitions and connections
- standard approach:
  - minimal approach + read the corresponding parts of [1] and [5]
- interested students:
  - standard approach + read [4]

students with problems w.r.t. graph theory: read [2]



Groups

- Where do groups of humans play a role in science?
  - computer science (teams in groupware, UNIX groups, etc.)
  - law science (groups as legal entities (GmbH, Ltd.))
  - economics (Working teams (project management), target groups for marketing, buyer groups etc.),
  - ethnology (ethnic groups & their characteristics),
  - history (e.g. social and political groups of the past & their role in historic societies),
  - art history (e.g. artist groups (Bauhaus, Brücke, Surrealists) with distinct philosophy, manifests & organizational frame)
  - sociology (obviously)

Groups in Social Psychology

- F. Tönnis (1887)[3]: Gemeinschaft ←→ Gesellschaft
- From 1930s: Small group research (see [4,5,6])
- Historically:
  Individualist school of thought (All phenomena and structures in a SN (incl groups) can be derived from analyzing dyadic individual relations)
  ←→Collectivistic school of thought (assign reality and parameters to groups independent of its members). Modern view : Emergence
- Homans (1950) [6]: "*A group is a number of persons who communicate with one another often over a span of time, and who are few enough so that each person is able to communicate with all the others, not at second hand, through other people, but face-to-face.*"

Groups in Social Psychology

- Number of group members < 20 (see [7])←→ human social perception limits)
- Group members: Share network of interpersonal attraction ([4, 5])
- Often: common goals, common norms, special communication structure, a special role- and affect-structure, group awareness ([4, 7])
- Small groups (e.g. friends clique) ←→ large groups (e.g. political party)
- Primary group (e.g. familiy) ←→ secondary group (e.g. colleagues)
- In-group ("my group") (special in group is reference group) ←→ out-group ("the others")
- Quasi groups (Profile clusters only)
- "Crowd", "mass", "clique", "gang", "community", "company", "squad", "team", ...

Groups in Social Psychology

- Number of group members < 20 (see [7])←→ human social perception limits)
- Group members: Share network of interpersonal attraction ([4, 5])
- Often: common goals, common norms, special communication structure, a special role- and affect-structure, group awareness ([4, 7])
- Small groups (e.g. friends clique) ←→ large groups (e.g. political party)
- Primary group (e.g. familiy) ←→ secondary group (e.g. colleagues)
- In-group ("my group") (special in group is reference group) ←→ out-group ("the others")
- Quasi groups (Profile clusters only)
- "Crowd", "mass", "clique", "gang", "community", "company", "squad", "team", ...

Groups in Social Psychology

- Number of group members < 20 (see [7])←→ human social perception limits)
- Group members: Share network of interpersonal attraction ([4, 5])
- Often: common goals, common norms, special communication structure, a special role- and affect-structure, group awareness ([4, 7])
- Small groups (e.g. friends clique) ←→ large groups (e.g. political party)
- Primary group (e.g. familiy) ←→ secondary group (e.g. colleagues)
- In-group ("my group") (special in group is reference group) ←→ out-group ("the others")
- Quasi groups (Profile clusters only)
- "Crowd", "mass", "clique", "gang", "community", "company", "squad", "team", ...

Finding Dense Subnetworks in Social Networks

Two basic possibilities to determine groups:

- Cluster profile elements of individuals (danger: quasi groups)
- or determine groups via social network (→ sociometry / network analysis)




Finding Dense Subnetworks in Social Networks

- What characterizes groups in sociometry? [11, 2]: groups are sub-graphs in a social network with the following properties:
  - Density: groups are denser than randomly chosen sub-graphs, (nodes have large neighborhood in G) → "Intra cluster coherence"
  - Compactness: mean average path-lengths are small within groups and/or connectivity is high (compare [1] for definitions) → "Intra cluster coherence"
  - Mutuality: many ties are reciprocal → "Intra cluster coherence"
  - Separation: group members have more ties within the group than outside → "inter cluster decoherence"
- Criteria are not independent: Moon [12]: Each member is connected to at least 1/k other members → distance between members is at most k. (see [2])

Finding Dense Subnetworks in Social Networks

- What characterizes groups in sociometry? [11, 2]: groups are sub-graphs in a social network with the following properties:
  - Density: groups are denser than randomly chosen sub-graphs, (nodes have large neighborhood in G) → "Intra cluster coherence"
  - Compactness: mean average path-lengths are small within groups and/or connectivity is high (compare [1] for definitions) → "Intra cluster coherence"
  - Mutuality: many ties are reciprocal → "Intra cluster coherence"
  - Separation: group members have more ties within the group than outside → "inter cluster decoherence"
- Criteria are not independent: Moon [12]: Each member is connected to at least 1/k other members → distance between members is at most k. (see [2])

Cliques

- A subset $U \subseteq V$ of a Graph (V,E) is a clique if G([U]) is a complete graph; G([U]) is the sub-graph induced by U.
- A clique is maximal if there is no clique U' with $U \subset U'$ in G
- A clique is a maximum clique if there is no clique with more vertices in G

Cliques

- Cliques are "perfect" in that they are
  - perfectly dense: Maximum degree Δ (G([U])) = |U|-1; minimum degree δ (G([U])) = |U|-1; average degree $\bar{\delta}$ (G([U]))=|U|-1
  - perfectly compact: diam(G([U]))=1, mean av. path length = 1, perfectly connected: if |U|=k then G([U]) is (k-1) vertex- and edge-connected

---

G is n-vertex connected if |V| > n and G – X is connected for every $X \subset V$ with |X| < n;

G is n-edge connected if |V| > 2 and and G – Y is connected for every $Y \subset E$ with |Y| < n;

Cliques

- Cliques are "perfect" in that they are
  - perfectly dense: Maximum degree Δ (G([U])) = |U|-1; minimum degree δ (G([U])) = |U|-1; average degree $\bar{\delta}$ (G([U]))=|U|-1
  - perfectly compact: diam(G([U]))=1, mean av. path length = 1, perfectly connected: if |U|=k then G([U]) is (k-1) vertex- and edge-connected

---

G is n-vertex connected if |V| > n and G – X is connected for every $X \subset V$ with |X| < n;

G is n-edge connected if |V| > 2 and and G – Y is connected for every $Y \subset E$ with |Y| < n;