2 Fundamentals

In this chapter we discuss basic terminology and notation for graphs, some fundamental algorithms, and a few other mathematical preliminaries.

We denote the set of integers by \( \mathbb{Z} \), the set of real numbers by \( \mathbb{R} \), and the set of rationals by \( \mathbb{Q} \). For a set \( X \) of numbers, \( X^+ \) denotes the subset of positive numbers in \( X \), and \( X^- \) the subset of non-positive numbers. The set of positive integers is denoted by \( \mathbb{Z}^+ \) and the set of non-negative integers by \( \mathbb{Z}^0 \).

We use \( \mathbb{R}^{m \times n} \) to denote the set of all real-valued matrices with \( m \) rows and \( n \) columns. If the entries of a matrix can be complex numbers, we write \( \mathbb{C}^{m \times n} \).

The \( n \)-dimensional identity matrix is denoted by \( I_n \). The \( n \)-dimensional vector with all entries equal to 1 (equal to 0) is denoted by \( \mathbf{1}_n \) (by \( \mathbf{0}_n \)).

For two functions \( f : \mathbb{R} \to \mathbb{R} \) and \( g : \mathbb{R} \to \mathbb{R} \), we say that \( f \) is in \( O(g) \) if there are positive constants \( c, c_0 \in \mathbb{R} \) such that \( f(x) \leq c \cdot g(x) \) holds for all \( x \geq c_0 \). Furthermore, we say that \( f \) is in \( O(g) \) if \( g \) is in \( O(f) \). This notation is useful to estimate the asymptotic growth of functions. In particular, running-times of algorithms are usually specified using this notation.

3 Centrality Indices

Centrality indices are to quantify an intuitive feeling that in most networks some vertices or edges are more central than others. Many vertex centrality indices were introduced for the first time in the 1950’s e.g., the Betweenness index [50, 51], degree centrality [40], or a first feedback centrality, introduced by Szebényi [52]. These early centrality indices also triggered a rash of research in which many new applications were found. However, not every centrality index is suitable to every application, so with time, dozens of new centrality indices were published. This chapter will present some of the more influential, classic centrality indices. We do not strive for completeness, but hope to give a catalog of basic centrality indices with some of their main applications.

In Section 3.1 we will begin with two simple examples to show how centrality indices can help to analyze networks and the situation these networks represent. In Section 3.2 we discuss the properties that are minimally required for a real-valued function on the set of vertices or edges of a graph to be a centrality index for vertices or edges, respectively. In subsequent Sections 3.3-3.9, various families of vertex and edge centralities are presented. First, centrality indices based on distance and neighborhood are discussed in Section 3.3. Additionally, this section presents in detail some instances of facility location problems as a possible application for centrality indices. Next we discuss the centrality indices based on shortest paths (see

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 mbf:
  - limit edge capacity to one
  - \( \text{mbf}(i) := \text{maximum possible flow through } i \text{ over all possible solutions to the s}-\text{maximum flow problem, averaged over all s and t.} \)

(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 $\text{rwb}$: idea:
  $\text{rwb}(i) := \text{number of times that a random walk starting at } s \text{ and ending at } t \text{ passes through } i \text{ along the way}, \text{ averaged over all } s \text{ and } t$

- $\text{rwb} \leftrightarrow \text{spb}$: opposite ends:
  - $\text{rwb}$: info has no idea where its going
  - $\text{spb}$: info knows exactly where its going

- compute for all $i$ $\text{rwb}(i)$: $O((m+n)n^2)$ worst case time complexity (comparable to $\text{spb}$)

Kirchhoff's Point Law (Current Conservation): total current flow in / out of node is zero:

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

- $A_{ij} = \begin{cases} 1 \\ 0 \end{cases}$ if there is an edge between $i$ and $j$, otherwise,
- $\delta_{ij} = \begin{cases} 1 \\ 0 \end{cases}$ if $i = j$, otherwise.
Kirchhoff's point law (current conservation): total current flow in / out of node is zero:

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

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

Kirchhoff's point law (current conservation): total current flow in / out of node is zero:

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

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

Kirchhoff's point law (current conservation): total current flow in / out of node is zero:

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

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

Kirchhoff's point law (current conservation): total current flow in / out of node is zero:

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

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

\[
\sum_j A_{ij} = k_i, \text{ the degree of vertex } i
\]

\[
\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \quad \rightarrow \quad (D - A) \cdot \mathbf{v} = \mathbf{s}
\]

“Graph Laplacian”

\( D \) is the diagonal matrix with elements \( D_{ii} = k_i \)

source vector \( \mathbf{s} \)

\[
s_i = \begin{cases} +1 & \text{for } i = s, \\ -1 & \text{for } i = t, \\ 0 & \text{otherwise}. \end{cases}
\]

\[
\mathbf{v} = (D - A)^{-1} \cdot \mathbf{s}
\]
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

\[
\sum_j A_{ij} = k_i, \text{ the degree of vertex } i
\]

\[
\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \quad \rightarrow \quad (\mathbf{D} - \mathbf{A}) \cdot \mathbf{V} = \mathbf{s}
\]

"Graph Laplacian"

\[
\mathbf{D} \text{ is the diagonal matrix with elements } D_{ii} = k_i
\]

source vector \( \mathbf{s} \)

\[
 s_i = \begin{cases} 
 +1 & \text{for } i = s, \\
 -1 & \text{for } i = t, \\
 0 & \text{otherwise.}
\end{cases}
\]

\[ \mathbf{V} = (\mathbf{D} - \mathbf{A})^{-1} \cdot \mathbf{s} \]

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

\[
\sum_j A_{ij} = k_i, \text{ the degree of vertex } i
\]

\[
\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it} \quad \rightarrow \quad (\mathbf{D} - \mathbf{A}) \cdot \mathbf{V} = \mathbf{s}
\]

"Graph Laplacian"

\[
\mathbf{D} \text{ is the diagonal matrix with elements } D_{ii} = k_i
\]

source vector \( \mathbf{s} \)

\[
 s_i = \begin{cases} 
 +1 & \text{for } i = s, \\
 -1 & \text{for } i = t, \\
 0 & \text{otherwise.}
\end{cases}
\]

\[ \mathbf{V} = (\mathbf{D} - \mathbf{A})^{-1} \cdot \mathbf{s} \]
(D - A) \cdot V = s

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

\( V = (D_v - A_v)^{-1} \cdot s \quad \text{(matrix inversion: } O(n^3)) \)

let us now add a \( v \)-th row and column back into \( (D_v - A_v)^{-1} \) with values all equal to zero.

The resulting matrix we will denote \( T \).

\[ V_i^{(st)} = T_{is} - T_{it} \]

\( \rightarrow \) 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])

\[(D - A) \cdot V = s\]

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

\[V = (D_v - A_v)^{-1} \cdot s\]  \(\text{matrix inversion: } O(n^3)\)

let us now add a \(v\)th row and column back into \((D_v - A_v)^{-1}\)
with values all equal to zero.
The resulting matrix we will denote \(T\).

\[V_i^{(st)} = T_{is} - T_{it}\]

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

Kirchhoff's 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])**

- Kirchhoff’s point law (current conservation): total current flow in/out of node is zero:
  \[
  \sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{st},
  \]
  \[A_{ij} = \begin{cases} 
  1 & \text{if there is an edge between } i \text{ and } j, \\
  0 & \text{otherwise},
  \end{cases}
  \]
  \[\delta_{ij} = \begin{cases} 
  1 & \text{if } i = j, \\
  0 & \text{otherwise}.
  \end{cases}
  \]

- 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 \(t\):
  \[
  I_s^{(st)} = 1, \quad I_t^{(st)} = 1.
  \]

- cfb(i) (denoted as \(b_i\)) is then:
  \[
  b_i = \frac{1}{2} \sum_{s \neq t} I_i^{(st)} \quad \text{(takes O(m n^2) for all } i \rightarrow \text{ for every node)}
  \]

**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 \(t\):
  \[
  I_s^{(st)} = 1, \quad I_t^{(st)} = 1.
  \]

- cfb(i) (denoted as \(b_i\)) is then:
  \[
  b_i = \frac{1}{2} \sum_{s \neq t} I_i^{(st)} \quad \text{(takes O(m n^2) for all } i \rightarrow \text{ for every node)}
  \]

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

- cfb == random walk betweenness centrality (rwb):
  \[
  b_i \\
  \]

- rwb(i): move around „messages“: start (absorbing) random walk at \(s\), end at \(t\):
  \[
  \text{rwb}(i) := \text{net number of times that a message passes through } i \text{ on its journey (averaged over a large number of trials and averaged over } s, t) \\
  \]

- if in node \(j\), probability that in next step at node \(i\) is:
  \[
  M_{ij} = \frac{A_{ij}}{k_j}, \quad \text{for } j \neq t, \\
  M = A \cdot D^{-1} \quad \text{with } D = \text{diag}(k_i) \\
  D_{ii} = k_i
  \]
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

\( cf_{\text{b}} == \text{random walk betweenness centrality (rwb)}: \)

\( \text{rwb}(i): \) move around „messages“: start (absorbing) random walk at \( s \), end at \( t \):

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

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

\( \text{if in node } j, \text{ probability that in next step at node } i \text{ is:} \)

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

\[
M = A \cdot D^{-1} \quad \text{with } D = \text{diag}(k)
\]

\[
D_{ii} = k_i
\]

we never leave \( t \), once we get there ("Hotel California effect“ :)) \( \rightarrow M_{it} = 0 \text{ for all } i \)

\( \rightarrow \text{possible: remove column } t \text{ without affecting transitions between any other vertices;} \)

denote by \( M_t = A_t \cdot D_t^{-1} \) the matrix with these elements removed, and similarly for \( A_t \) and \( D_t \).

\( \text{for a walk starting at } s, \text{ the probability that we find ourselves at vertex } j \text{ after } r \text{ steps is given by} \)

\[
[M^r]_{js}
\]

\( \text{probability that we then take a step to an adjacent vertex } i \text{ is} \)

\[
k_j^{-1}[M^r]_{js}
\]
Random Walk Centrality == Current Flow Btw. Centrality (see [5])

- we never leave $t$, once we get there ("Hotel California effect" :-() $\rightarrow M_{it} = 0$ for all $i$

$\rightarrow$ possible: remove column $t$ without affecting transitions between any other vertices;

denote by $M_t = A_t \cdot D_t^{-1}$ 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 $[M_t^r]_{js}$

- probability that we then take a step to an adjacent vertex $i$ is $k_j^{-1}[M_t^r]_{js}$.

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

- previous slide: probability at $j$ after $r$ steps and then $j \rightarrow i$ was:

$$k_j^{-1}[M_t^r]_{js}$$

- summing over $r$ from 0 to $\infty$ $\rightarrow$ geometric series $\rightarrow$

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

$$k_j^{-1}[(I - M_t)^{-1}]_{js}$$

$\rightarrow$ $V = D_t^{-1} \cdot (I - M_t)^{-1} \cdot s = (D_t - A_t)^{-1} \cdot s.$

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

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 $j \rightarrow i$ was:

$$k_j^{-1}[M_t^r]_{js}$$

- summing over $r$ from 0 to $\infty$ $\rightarrow$ geometric series $\rightarrow$

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

$$k_j^{-1}[(I - M_t)^{-1}]_{js}$$

$\rightarrow$ $V = D_t^{-1} \cdot (I - M_t)^{-1} \cdot s = (D_t - A_t)^{-1} \cdot s.$

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

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:

\[ k_j^{-1} [M_r]_{js}. \]

- summing over r from 0 to \( \infty \): \( \rightarrow \) geometric series \( \rightarrow \)

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

\[ k_j^{-1} \cdot (I - M) \cdot s = (D_t - A_t)^{-1} \cdot s. \]

\[
V = D_t^{-1} \cdot (I - M_t)^{-1} \cdot s \quad \text{(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

Example ([5])

Network 1

<table>
<thead>
<tr>
<th>network</th>
<th>betweenness measure</th>
<th>shortest-path</th>
<th>max-flow</th>
<th>random walk / current-flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network 1:</td>
<td>vertices A &amp; B</td>
<td>0.636</td>
<td>0.631</td>
<td>0.670</td>
</tr>
<tr>
<td>vertex C</td>
<td>0.200</td>
<td>0.282</td>
<td>0.333</td>
<td></td>
</tr>
<tr>
<td>vertices X &amp; Y</td>
<td>0.200</td>
<td>0.068</td>
<td>0.269</td>
<td></td>
</tr>
<tr>
<td>Network 2:</td>
<td>vertices A &amp; B</td>
<td>0.265</td>
<td>0.269</td>
<td>0.321</td>
</tr>
<tr>
<td>vertex C</td>
<td>0.243</td>
<td>0.004</td>
<td>0.267</td>
<td></td>
</tr>
<tr>
<td>vertices X &amp; Y</td>
<td>0.125</td>
<td>0.024</td>
<td>0.194</td>
<td></td>
</tr>
</tbody>
</table>

Network 2

Example ([5])

Network 1

<table>
<thead>
<tr>
<th>network</th>
<th>betweenness measure</th>
<th>shortest-path</th>
<th>max-flow</th>
<th>random walk / current-flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network 1:</td>
<td>vertices A &amp; B</td>
<td>0.636</td>
<td>0.631</td>
<td>0.670</td>
</tr>
<tr>
<td>vertex C</td>
<td>0.200</td>
<td>0.282</td>
<td>0.333</td>
<td></td>
</tr>
<tr>
<td>vertices X &amp; Y</td>
<td>0.200</td>
<td>0.068</td>
<td>0.269</td>
<td></td>
</tr>
<tr>
<td>Network 2:</td>
<td>vertices A &amp; B</td>
<td>0.265</td>
<td>0.269</td>
<td>0.321</td>
</tr>
<tr>
<td>vertex C</td>
<td>0.243</td>
<td>0.004</td>
<td>0.267</td>
<td></td>
</tr>
<tr>
<td>vertices X &amp; Y</td>
<td>0.125</td>
<td>0.024</td>
<td>0.194</td>
<td></td>
</tr>
</tbody>
</table>
**Example ([5])**

![Network 1](image1) ![Network 2](image2)

<table>
<thead>
<tr>
<th>network</th>
<th>betweenness measure</th>
<th>random walk / current-flow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>shortest-path</td>
<td>max-flow</td>
</tr>
<tr>
<td>Network 1:</td>
<td>vertices A &amp; B</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>vertex C</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>vertices X &amp; Y</td>
<td>0.200</td>
</tr>
<tr>
<td>Network 2:</td>
<td>vertices A &amp; B</td>
<td>0.265</td>
</tr>
<tr>
<td></td>
<td>vertex C</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>vertices X &amp; Y</td>
<td>0.125</td>
</tr>
</tbody>
</table>

**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 $W$
- centrality $s(v)$ of node $v$ is proportional to sum of centralities $s(w)$ of adjacent nodes $w$ (multiplied with corresponding edge weight). 
  
  centrality vector $s$ of the nodes is thus an *eigenvector* of $W$: $s=\lambda s$

In order to make this equation solvable, introduce a „centrality input“ or „external information“ $E(v)$ for every node $v$: $s=E+Ws$

$I-W$ is invertible if \( \sum_{k=0}^{\infty} W^k \) converges \( \iff \) the largest eigenvalue of $W$ is less than one (see [1]).
**Random Walk Centrality \( \approx \) Current Flow Btw. Centrality (see [5])

- previous slide: probability at \( j \) after \( r \) steps and then \( j \rightarrow i \) was:
  \[
  k_j^{-1} [M^r]_{js}.
  \]
- summing over \( r \) from 0 to \( \infty \) → geometric series →
  \[
  \sum_{r=0}^{\infty} M^r = (I - M)^{-1} \quad \text{if} \forall i: |\lambda_i| < 1 \quad \text{where} \ \lambda_i \ \text{Eigenvalues of} \ M
  \]
  \[
  j \rightarrow i \quad \text{with} \quad \sum_{r=0}^{\infty} M^r \cdot s_j = \lambda_i s_i
  \]
  \[
  \rightarrow \quad V = D_t^{-1} \cdot (I - M_t)^{-1} \cdot s = (D_t - A_t) \lambda^{-1} \cdot s.
  \]

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

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

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.

- Further example: Random surfer on Web-pages
- Directed unweighted graph \( G=(V,E) \)
- Define Markov transition matrix as
  \[
  t_{ij} = \begin{cases} 
  \frac{1}{\operatorname{deg}^+(i)} & \text{if } (i,j) \in E \\
  0 & \text{if } (i,j) \notin E \\
  \frac{1}{|V|} & \text{if } \operatorname{deg}^+(i) = 0
  \end{cases}
  \]

(choose one outgoing link randomly, probability inverse proportional to out degree of current node; if node is a sink (no outgoing links) choose a random page)
Question: is there a unique stationary distribution $\pi$? (⇒ in essence is the chain irreducible and positively recurrent?)

→ make it irreducible: $T = \alpha T + (1- \alpha)E$ 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:

$$t_y = \frac{A_y}{\text{deg}(i)} \Rightarrow \pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$$

Proof: $(\pi T)_y = \sum_{v \in V} \pi_v t_y = \sum_{v \in V} \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)} \frac{\sum \text{deg}(i)}{\sum_{v \in V} \text{deg}(v)} = \pi_j$

$\pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$

$\pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$

$\pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$

$\pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$

$\pi_i = \frac{\text{deg}(i)}{\sum_{v \in V} \text{deg}(v)}$
**Feedback-Centrality**

- **Question:** is there a unique stationary distribution \( \pi \)? (\( \Rightarrow \) in essence is the chain irreducible and positively recurrent?)

- make it irreducible: \( T = \alpha T + (1 - \alpha)E \) 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 \( \leftrightarrow \) degree centrality:** Assume undirected, unweighted graph with adjacency matrix \( A \); we have then:
  \[
  \pi_y = \frac{A_y}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v \in V} \deg(v)}
  \]

  Proof:
  \[
  (\pi T)_y = \sum_{v \in V} \frac{\deg(i)}{\sum_{v \in V} \deg(v)} \pi_y \frac{\sum_{v \in V} A_{yv}}{\sum_{v \in V} \deg(v)} = \frac{\deg(j)}{\sum_{v \in V} \deg(v)} = \pi_j
  \]

**Feedback-Centrality: Page Rank**

- **Famous ingredient of Google**

- Centrality of a web-page depends on the centralities of the pages linking to it:
  \[
  c(p) = d \sum_{q \in \text{in-neighbor of } p} \frac{c(q)}{\deg^+(q)} + (1 - d)
  \]
  where \( d \) is a damping factor; \( \deg^+(q) \) is the out degree of \( q \).

- Matrix Notation:
  \[
  \mathbf{c} = d \mathbf{Pc} + (1 - d)(1,1,...,1)^T
  \]
  where transition matrix \( P_{ij} = 1/\deg^+(j) \) if \( (j,i) \in E \) and \( P_{ij} = 0 \) otherwise.

- **Famous ingredient of Google**

- Centrality of a web-page depends on the centralities of the pages linking to it:
  \[
  c(p) = d \sum_{q \in \text{in-neighbor of } p} \frac{c(q)}{\deg^+(q)} + (1 - d)
  \]
  where \( d \) is a damping factor; \( \deg^+(q) \) is the out degree of \( q \).

- Matrix Notation:
  \[
  \mathbf{c} = d \mathbf{Pc} + (1 - d)(1,1,...,1)^T
  \]
  where transition matrix \( P_{ij} = 1/\deg^+(j) \) if \( (j,i) \in E \) and \( P_{ij} = 0 \) otherwise.
Solving the equation \( \mathbf{c} = d \mathbf{P} \mathbf{c} + (1 - d)(1,1,\ldots,1)^T \):

- If \(0 \leq d < 1\) the equation has a unique solution

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

- How do we compute the solution avoiding matrix inversion? \(\rightarrow\) Jacobi power iteration:

\[
c^{(k+1)}_i = d \sum_j P_{ij} c^{(k)}_j + (1 - d)
\]

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

\[
c^{(k+1)}_i = d \left( \sum_{j<i} P_{ij} c^{(k+1)}_j + \sum_{j \geq i} P_{ij} c^{(k)}_j \right) + (1 - d)
\]

Solving the equation \( \mathbf{c} = d \mathbf{P} \mathbf{c} + (1 - d)(1,1,\ldots,1)^T \):

- If \(0 \leq d < 1\) the equation has a unique solution

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

- How do we compute the solution avoiding matrix inversion? \(\rightarrow\) Jacobi power iteration:

\[
c^{(k+1)}_i = d \sum_j P_{ij} c^{(k)}_j + (1 - d)
\]

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

\[
c^{(k+1)}_i = d \left( \sum_{j<i} P_{ij} c^{(k+1)}_j + \sum_{j \geq i} P_{ij} c^{(k)}_j \right) + (1 - d)
\]
- **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]

---

- **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)
  - Collectivist 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. family) ↔ 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”, ....
Two basic possibilities to determine groups:

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

- 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⊆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 ⊆ U' in G.
  
  - A clique is a **maximum** clique if there is no clique with more vertices in G.
Clique are "perfect" in that they are

- perfectly dense: Maximum degree $\Delta(G([U])) = |U|-1$; minimum degree $\delta(G([U])) = |U|-1$; average degree $\bar{\delta}(G([U])) = |U|-1$
- perfectly compact: $\text{diam}(G([U])) = 1$, 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 $X \subseteq V$ with $|X| < n$;
$G$ is $n$-edge connected if $|V| > 2$ and and $G - Y$ is connected for every $Y \subseteq E$ with $|Y| < n$;

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