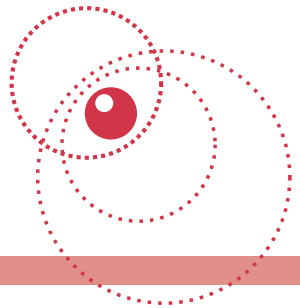


Web search

Graph Mining

7 July 2011

Pierre Senellart

TELECOM ParisTech

Outline

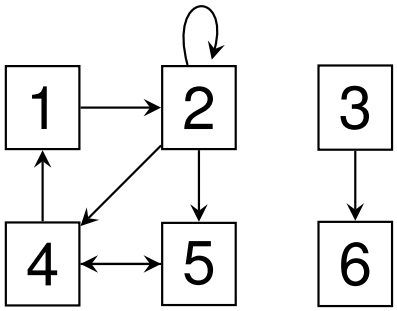
Basics of Graph Theory

The Web as a graph

Social Networking Sites

Conclusion

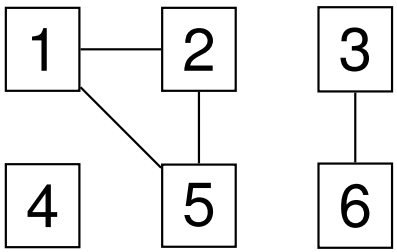
Graphs



Definition

A directed graph is a pair (S, A) where:

- S is a finite set of vertices (or nodes)
- A is a subset of S^2 defining the edges (or arcs)



Definition

An undirected graph is a pair (S, A) where:

- S is a finite set of vertices (or nodes)
- A is a set of (unordered) pairs of elements of S defining the edges (or arcs)

Remark

Graph is the mathematical term, network is used to describe real-world graphs.

Paths and Connectedness

Definition

A path is a sequence of vertices $v_1 \ldots v_n$ such that v_k is connected by an edge to v_{k+1} for $1 \leq k \leq n-1$.

Definition

The underlying undirected graph of a directed graph G is the graph obtained by adding all reverse edges.

Definition

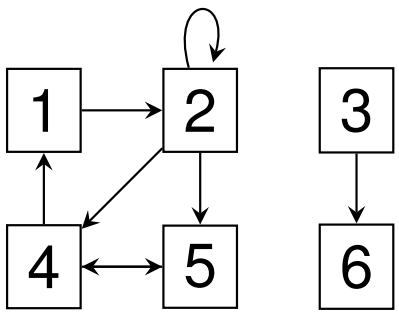
An undirected graph is connected if for every two vertices u and v, there exists a path starting from u and ending in v.
A directed graph is strongly connected if it is connected, and is weakly connected if the underlying undirected graph is connected.

Connected Components

Definition

(S', A') is a subgraph of (S, A) if $S' \subseteq S$ and A' is the restriction of A to edges whose vertices are in S'.

- Connected component: maximal connected subgraph
- Strongly connected component: maximal strongly connected subgraph
- Weakly connected component: maximal weakly connected subgraph

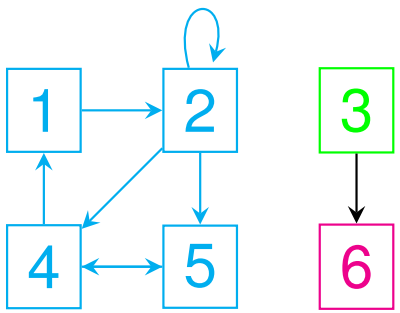


Connected Components

Definition

(S', A') is a subgraph of (S, A) if $S' \subseteq S$ and A' is the restriction of A to edges whose vertices are in S'.

- Connected component: maximal connected subgraph
- Strongly connected component: maximal strongly connected subgraph
- Weakly connected component: maximal weakly connected subgraph



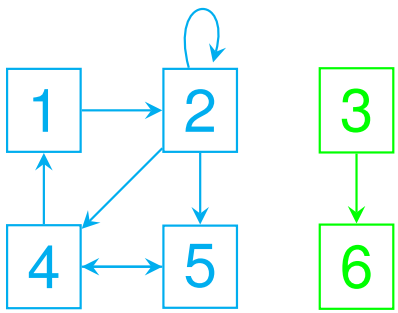
Strongly connected components

Connected Components

Definition

(S', A') is a subgraph of (S, A) if $S' \subseteq S$ and A' is the restriction of A to edges whose vertices are in S'.

- Connected component: maximal connected subgraph
- Strongly connected component: maximal strongly connected subgraph
- Weakly connected component: maximal weakly connected subgraph



Weakly connected components

Vocabulary

Incident: an edge is said to be incident to a vertex if it it has this vertex for endpoint

Degree (of a vertex): number of edges incident to a vertex, in an undirected graph

Indegree (of a vertex): number of edges arriving to a vertex, in a directed graph

Outdegree (of a vertex): number of edges leaving from a vertex, in a directed graph

Cycle: Path whose start and end vertex is the same

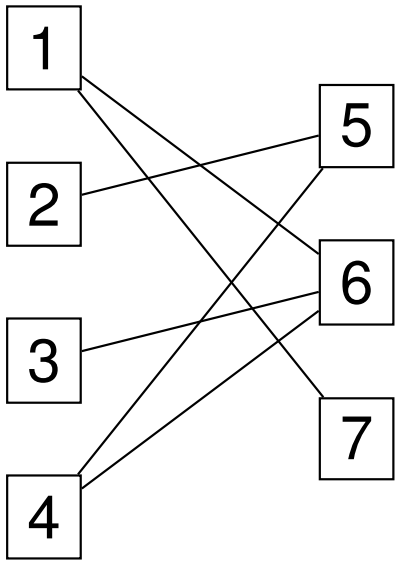
Distance: Length of the shortest path between two vertices

Sparse: a graph (S, A) is sparse if $|A| \ll |S|^2$

Bipartite graphs

Definition

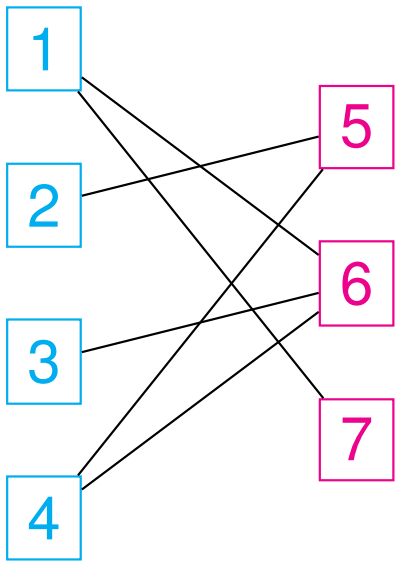
A bipartite graph is an undirected graph (S, A) such that $S = S_1 \cup S_2$ (with $S_1 \cap S_2 = \varnothing$), and no edge of A is incident to two vertices in S_1 or two vertices in S_2.



Bipartite graphs

Definition

A bipartite graph is an undirected graph (S, A) such that $S = S_1 \cup S_2$ (with $S_1 \cap S_2 = \varnothing$), and no edge of A is incident to two vertices in S_1 or two vertices in S_2.

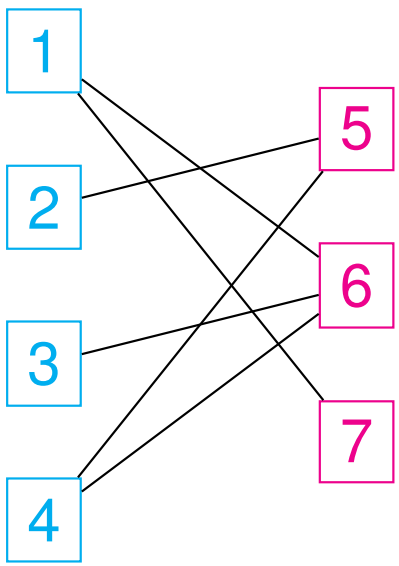


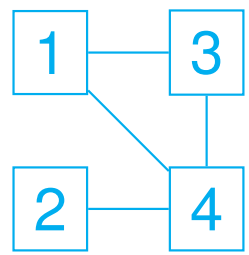
Bipartite graphs

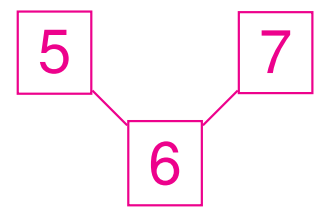
Definition

A bipartite graph is an undirected graph (S, A) such that $S = S_1 \cup S_2$ (with $S_1 \cap S_2 = \varnothing$), and no edge of A is incident to two vertices in S_1 or two vertices in S_2.

Paths of length 2 in a bipartite graph define two regular undirected graphs.

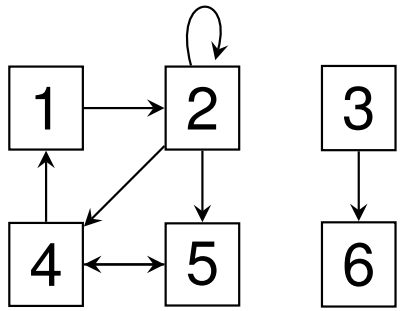






Matrix Representation of a Graph

A graph G can be represented by its adjacency matrix M:



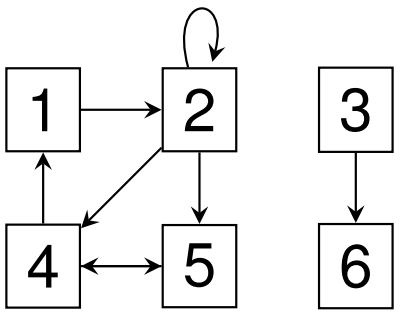
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- Adjacency matrices of undirected graphs are symmetric.
- M^{T} is the adjacency matrix obtained from G by reversing the arrows.
- M^n is the matrix of the graph of all paths of length n in G.

Concrete Representation of Graphs

In programming, graphs are usually represented as:

- its adjacency matrix (stored as a multidimensional array), for non-sparse graphs
- the list of all edges incident to each node, for sparse graphs



1	2
2	2, 4, 5
3	6
4	1, 5
5	4
6	

Outline

Basics of Graph Theory

The Web as a graph
- Introduction
- Computing the importance of a page
- Spamdexing
- Discovering communities

Social Networking Sites

Conclusion

Outline

Basics of Graph Theory

The Web as a graph
- Introduction
- Computing the importance of a page
- Spamdexing
- Discovering communities

Social Networking Sites

Conclusion

The Web Graph

The World Wide Web seen as a (directed) graph:

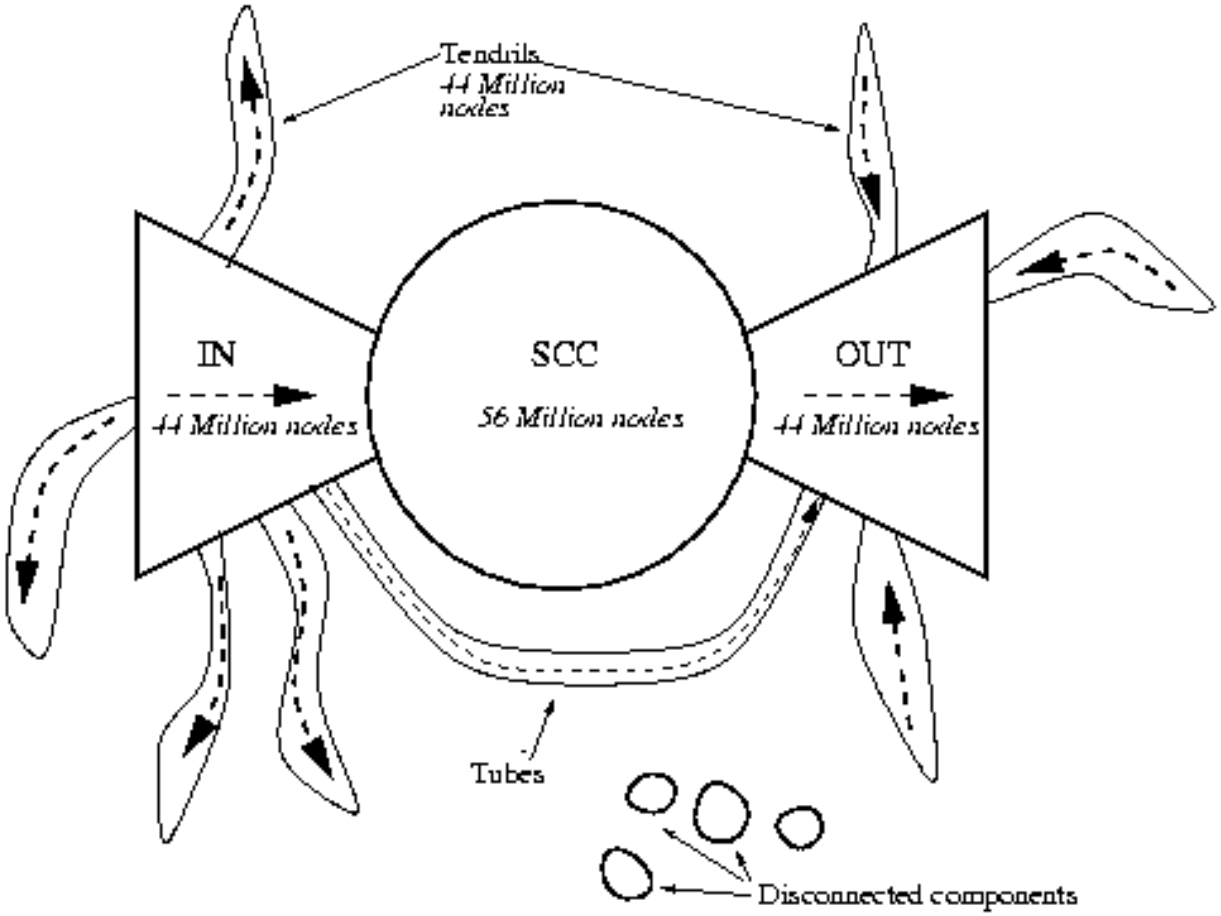
Vertices: Web pages

Edges: hyperlinks

Same for other interlinked environments:

- dictionaries
- encyclopedias
- scientific publications
- social networks

The shape of the Web graph



[Broder et al., 2000]

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?

Yes, the World Wide Web is sparse

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices?

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices? Logarithmic typical distance

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices? Logarithmic typical distance

Local clustering. If a is connected to both b and c, is the probability that b is connected to c significantly greater than the probability any two nodes are connected?

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices? Logarithmic typical distance

Local clustering. If a is connected to both b and c, is the probability that b is connected to c significantly greater than the probability any two nodes are connected? Strong local clustering

Characteristics of Interest of a Graph

Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices? Logarithmic typical distance

Local clustering. If a is connected to both b and c, is the probability that b is connected to c significantly greater than the probability any two nodes are connected? Strong local clustering

Degree distribution. What is the distribution of the degree of vertices?

Characteristics of Interest of a Graph

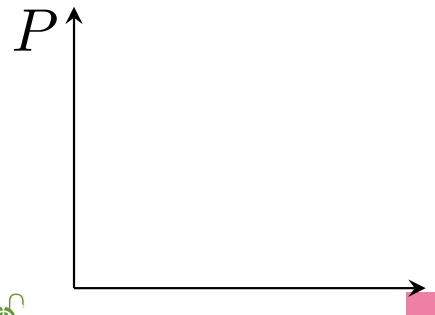
Sparsity. Is the graph sparse ($|A| \ll |S|^2$)?
Yes, the World Wide Web is sparse

Typical distance. What is the mean distance between any pairs of vertices? Logarithmic typical distance

Local clustering. If a is connected to both b and c, is the probability that b is connected to c significantly greater than the probability any two nodes are connected? Strong local clustering

Degree distribution. What is the distribution of the degree of vertices?

Power-law indegree and outdegree distribution ($2 \leq \gamma \leq 3$) [Broder et al., 2000]



Outline

Basics of Graph Theory

The Web as a graph
- Introduction
- Computing the importance of a page
- Spamdexing
- Discovering communities

Social Networking Sites

Conclusion

PageRank (Google's Ranking [Brin and Page, 1998])

Idea

Important pages are pages pointed to by important pages.

$$\begin{cases} g_{ij} = 0 & \text{if there is no link between page } i \text{ and } j; \\ g_{ij} = \frac{1}{n_i} & \text{otherwise, with } n_i \text{ the number of outgoing links of page } i. \end{cases}$$

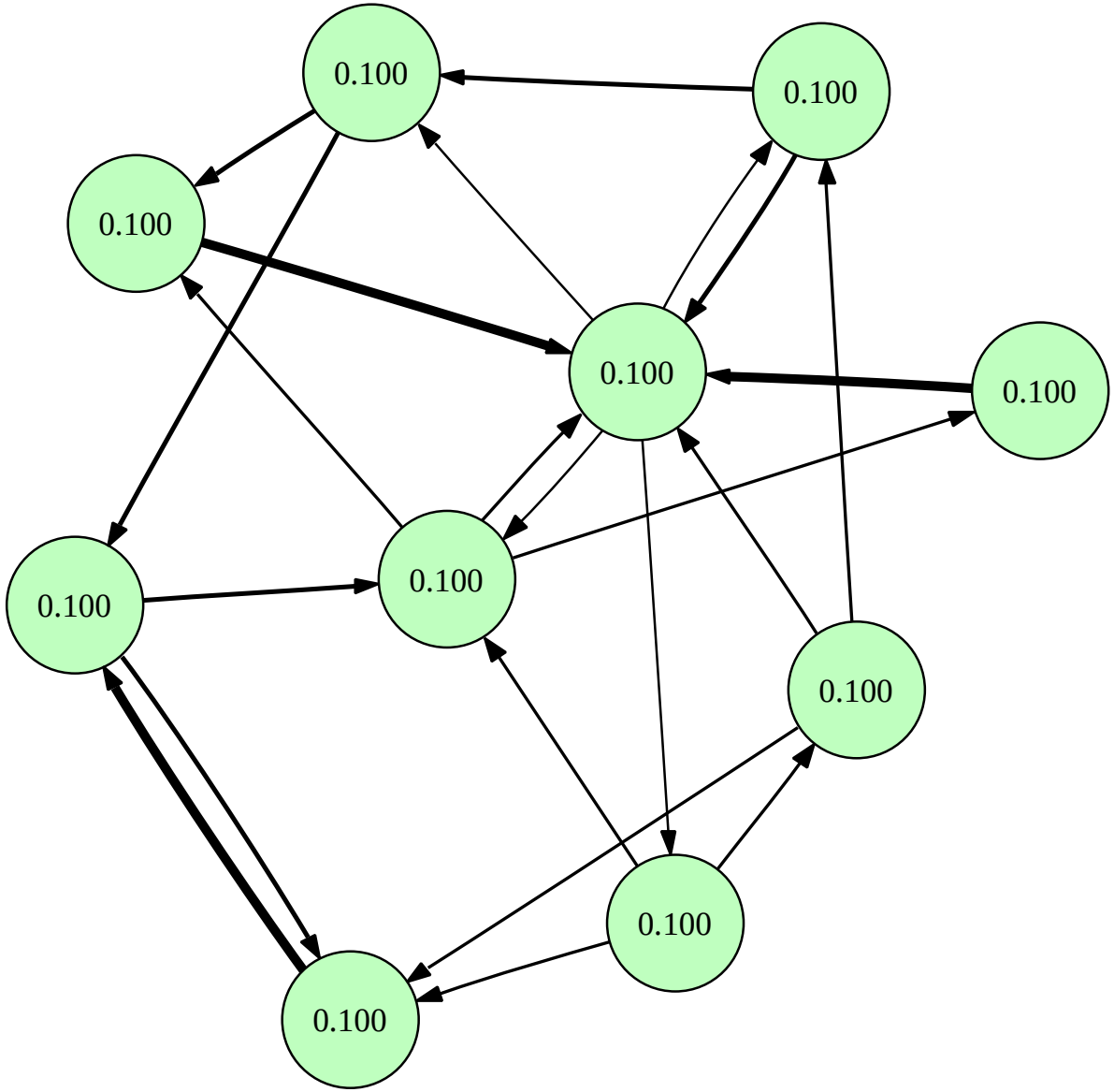
Definition (Tentative)

Probability that the surfer following the random walk in G has arrived on page i at some distant given point in the future.

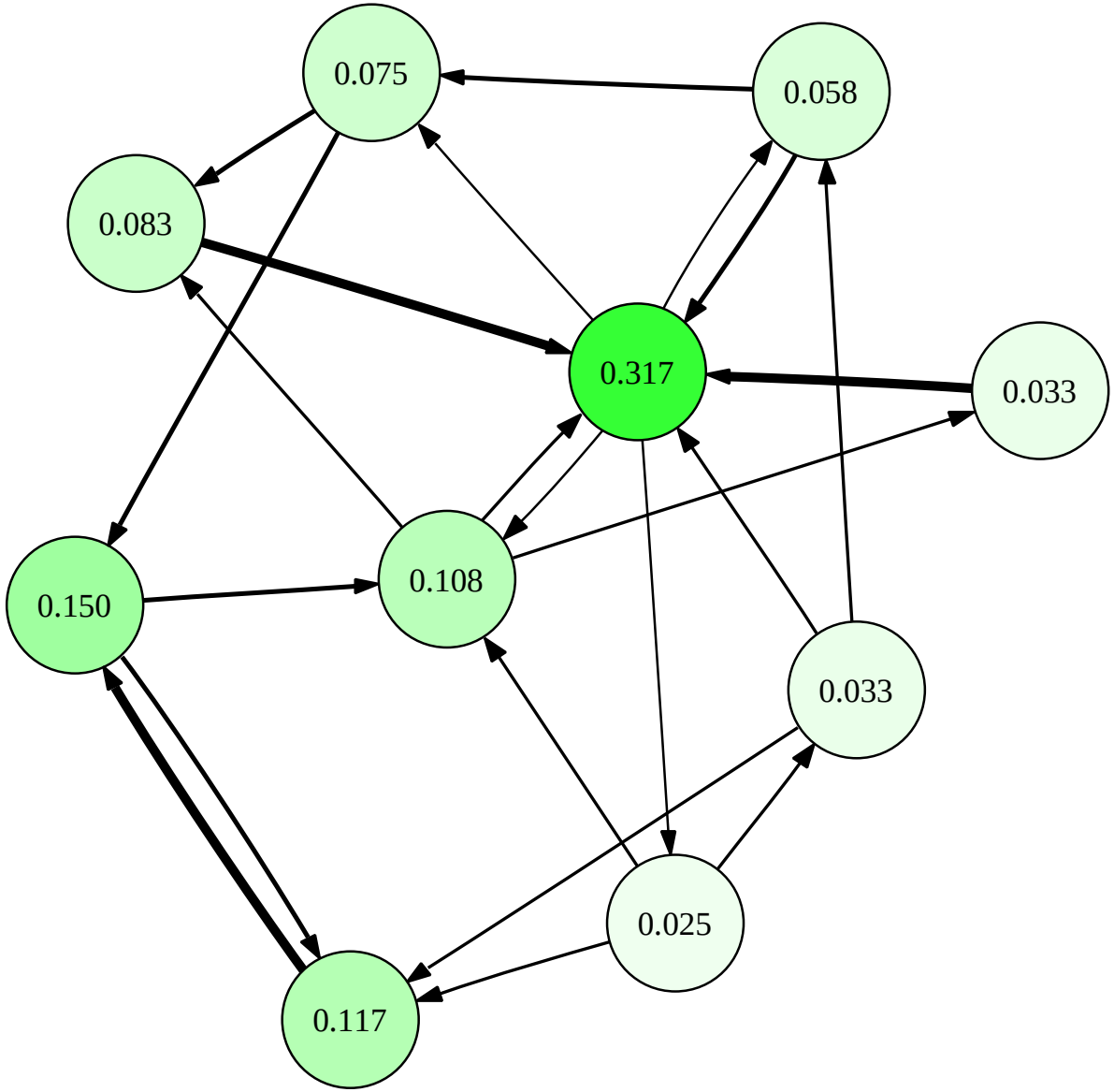
$$\mathrm{pr}(i) = \left(\lim_{k \to +\infty} (G^T)^k v \right)_i$$

where v is some initial column vector.

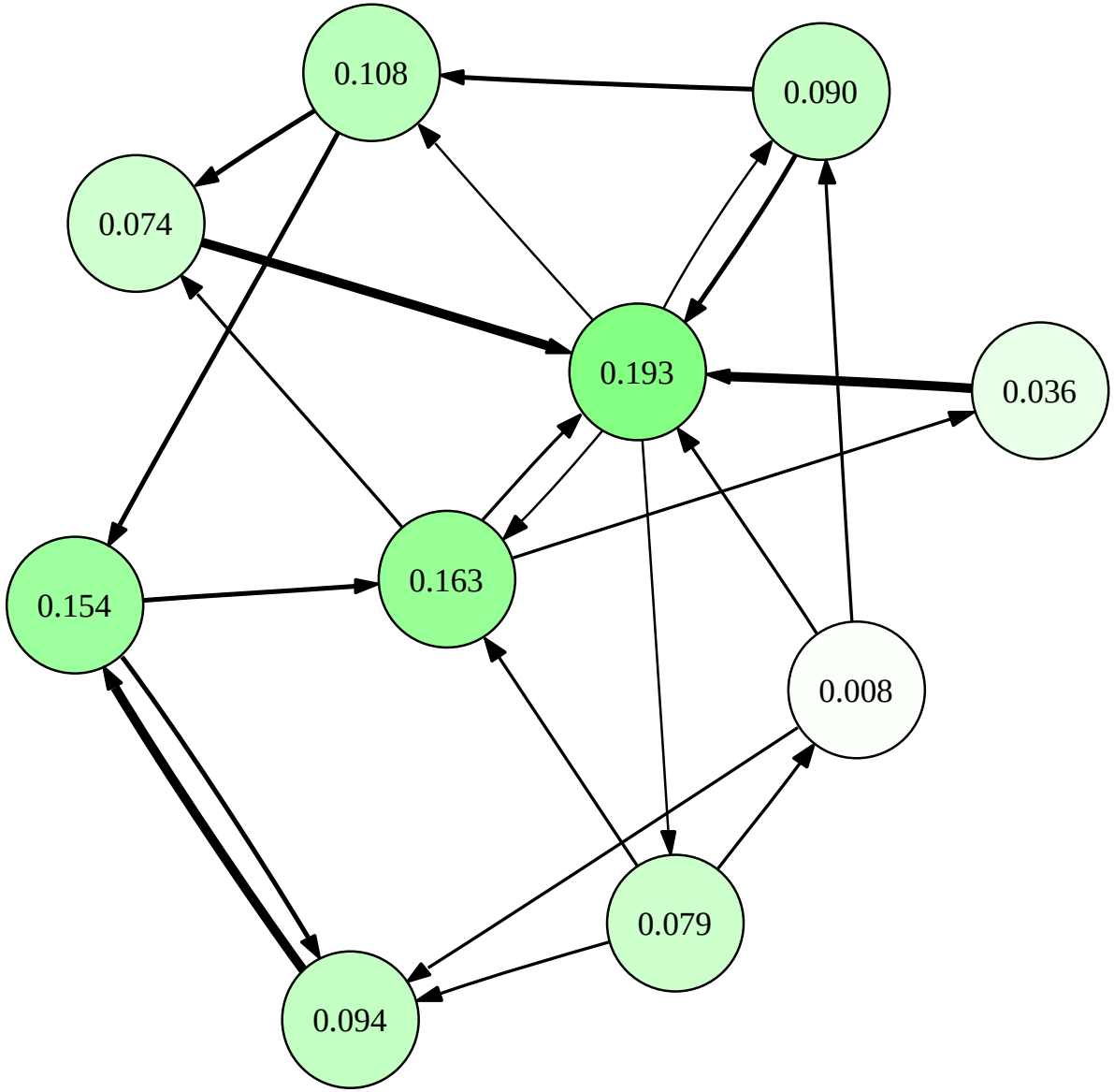
Illustrating PageRank Computation



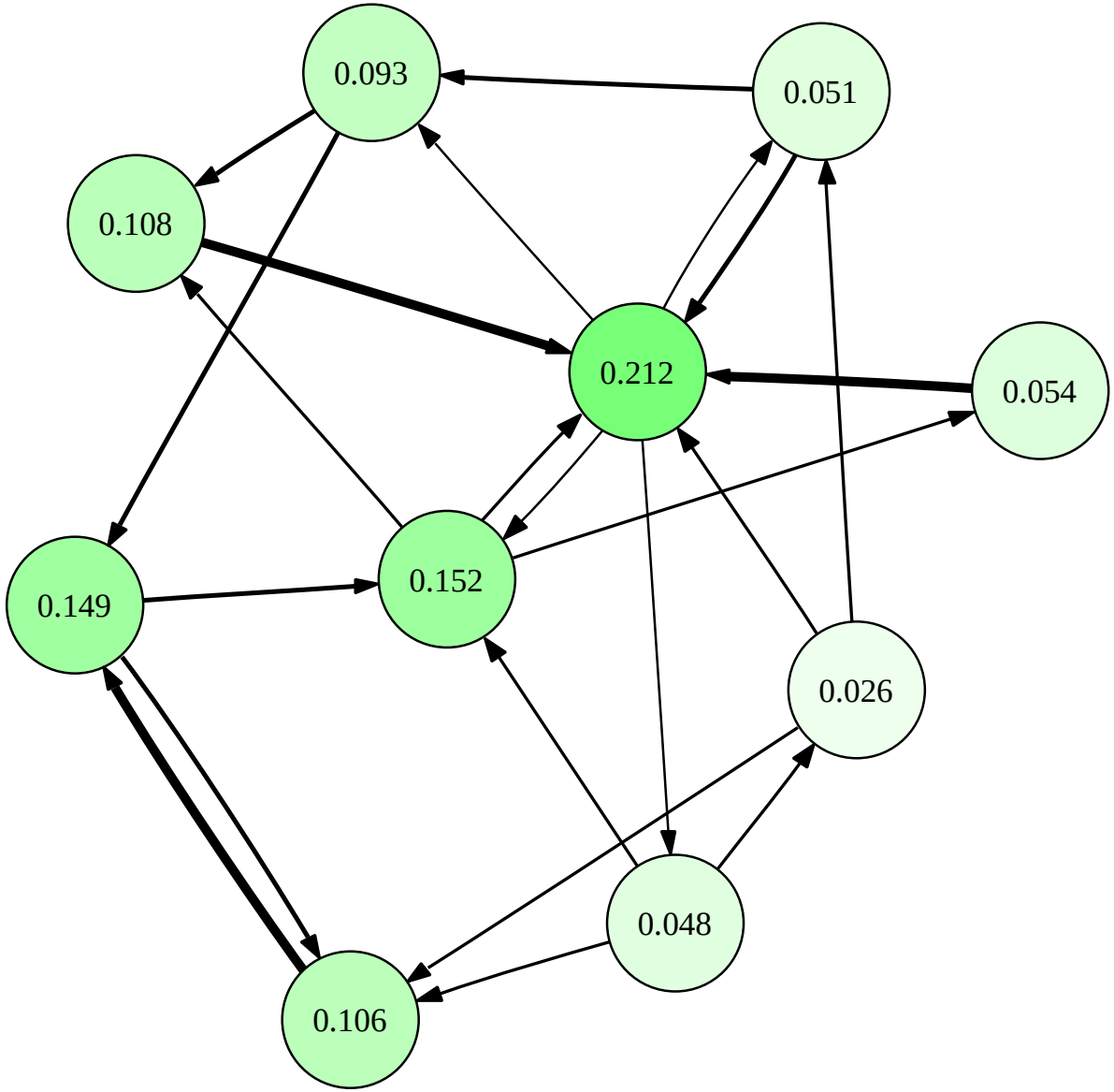
Illustrating PageRank Computation



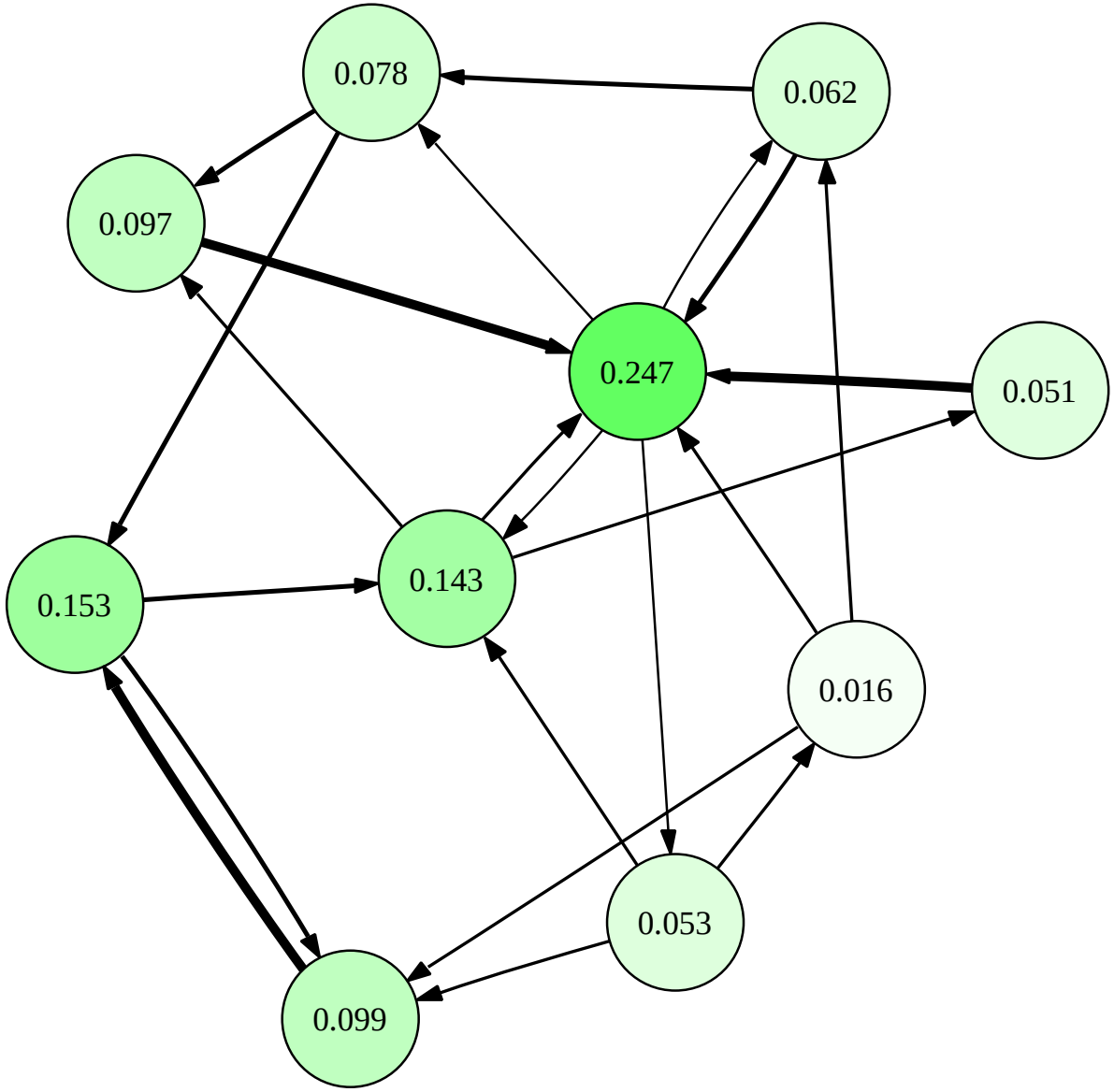
Illustrating PageRank Computation



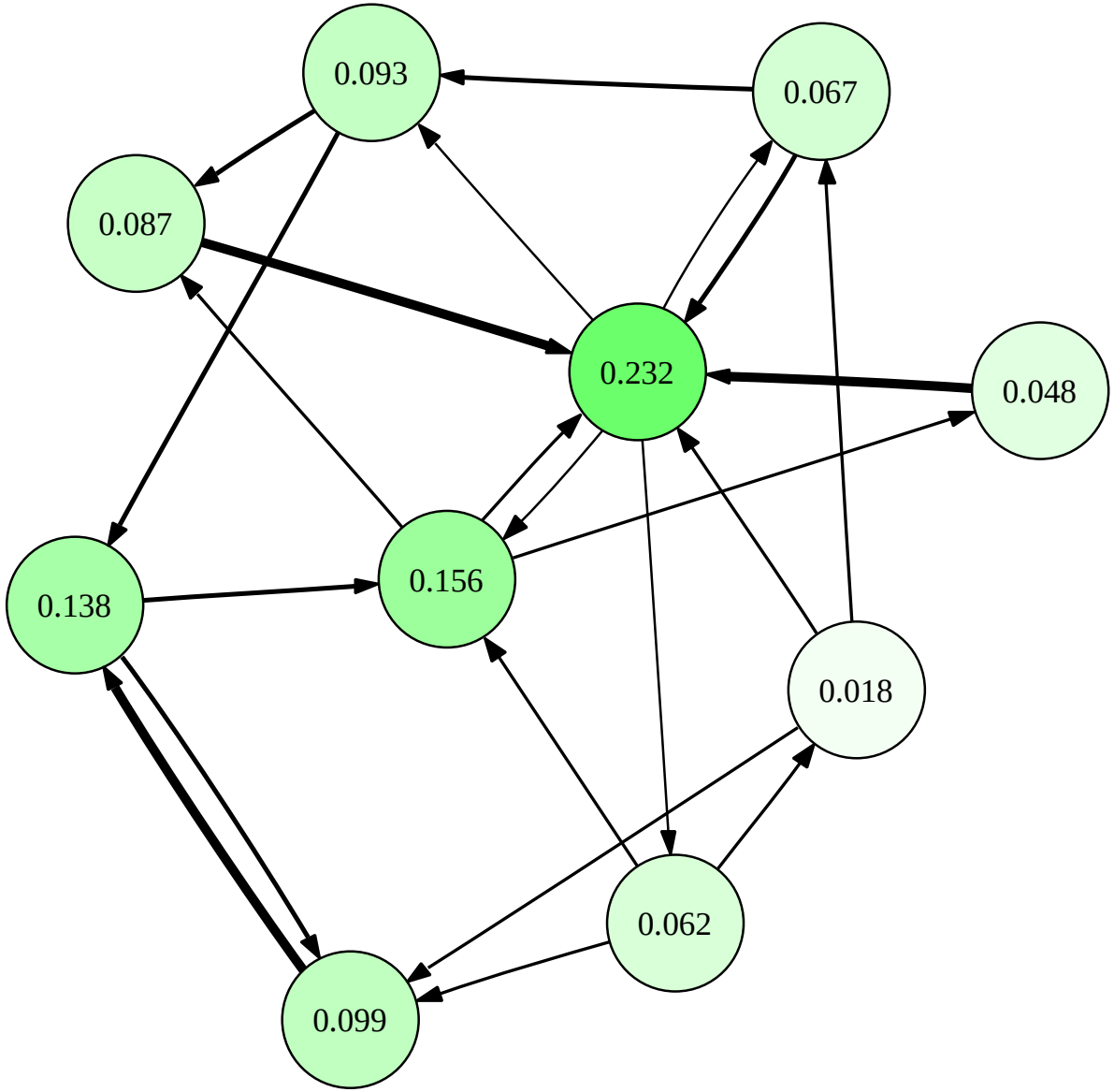
Illustrating PageRank Computation



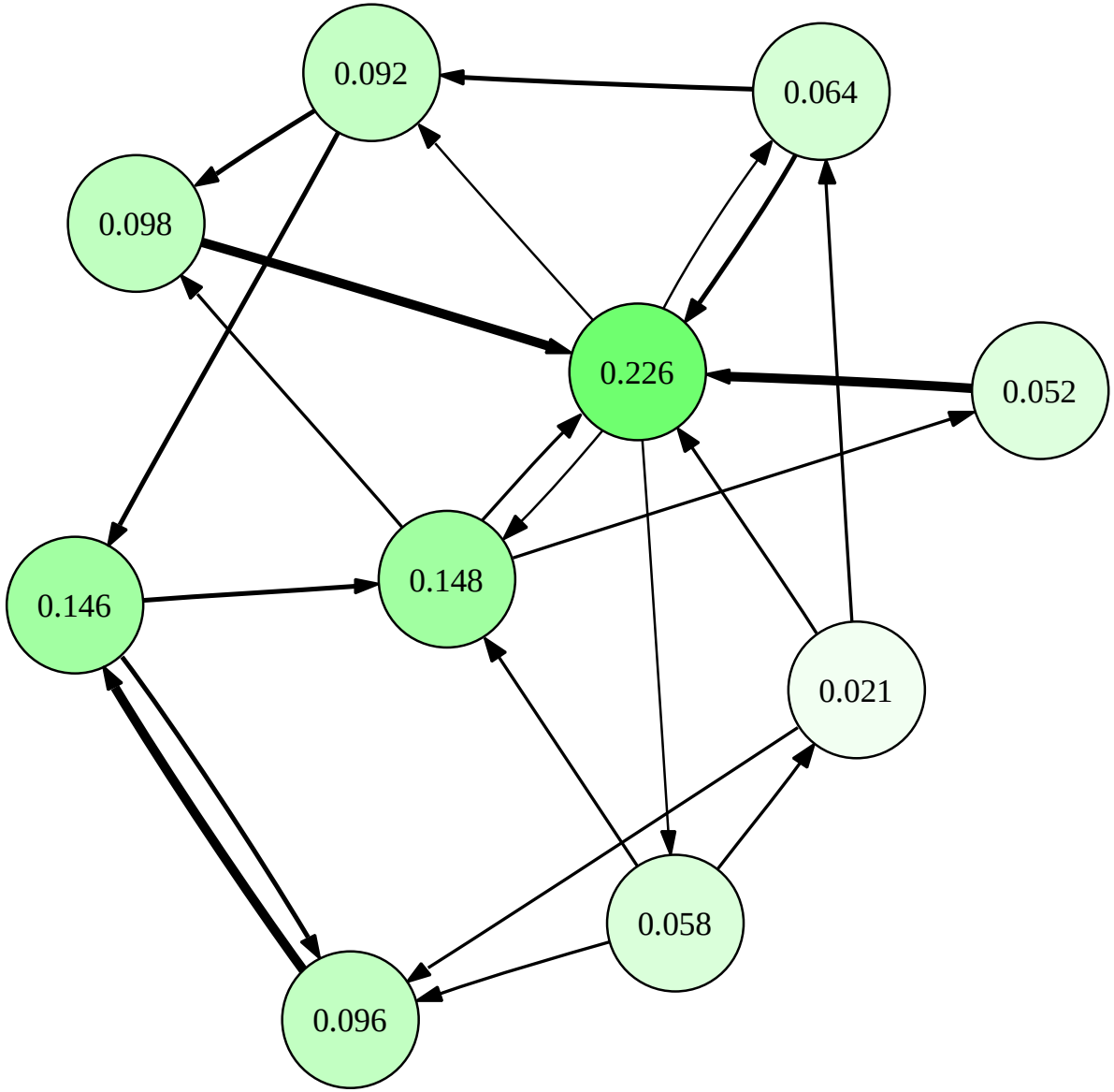
Illustrating PageRank Computation



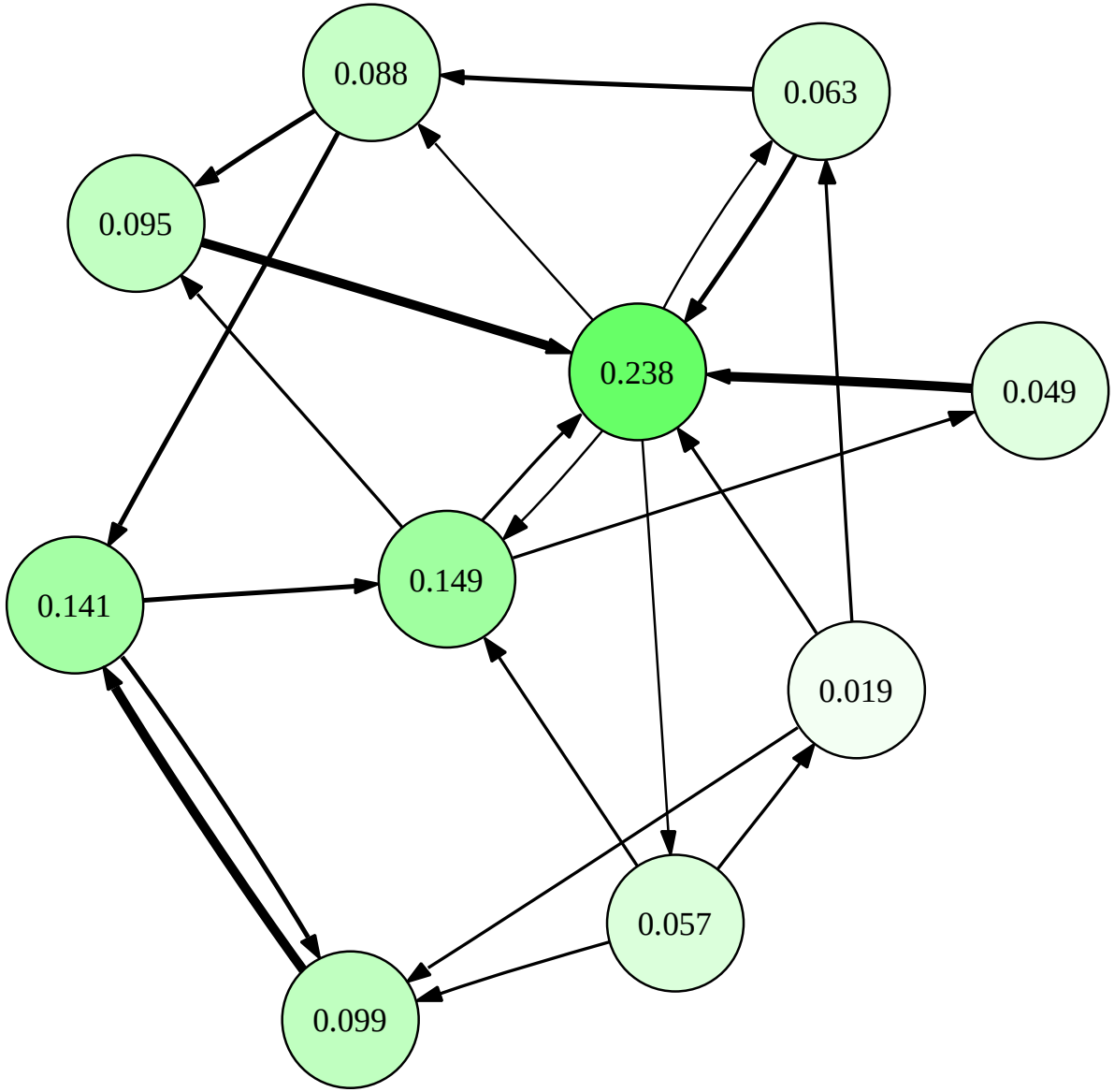
Illustrating PageRank Computation



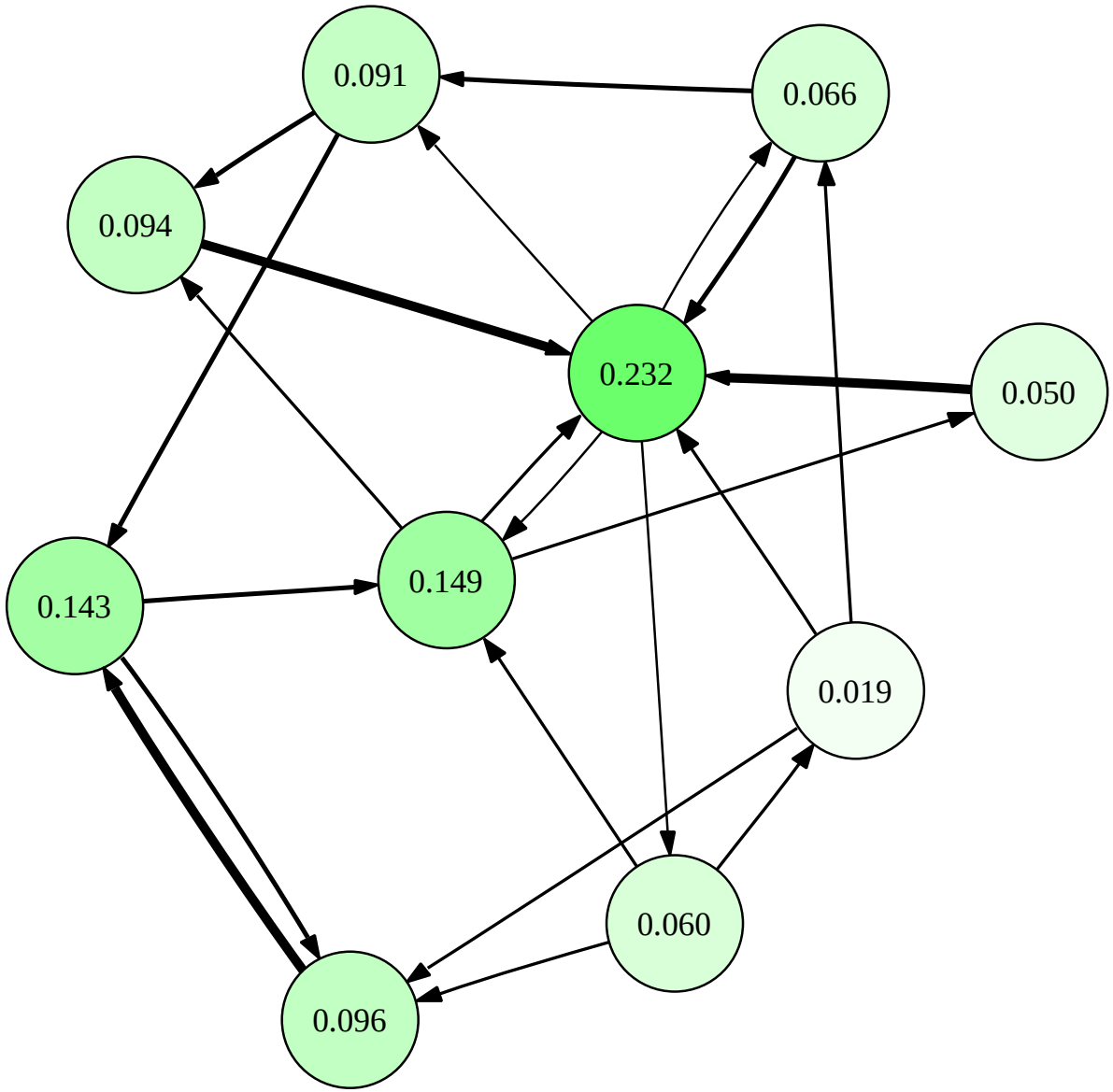
Illustrating PageRank Computation



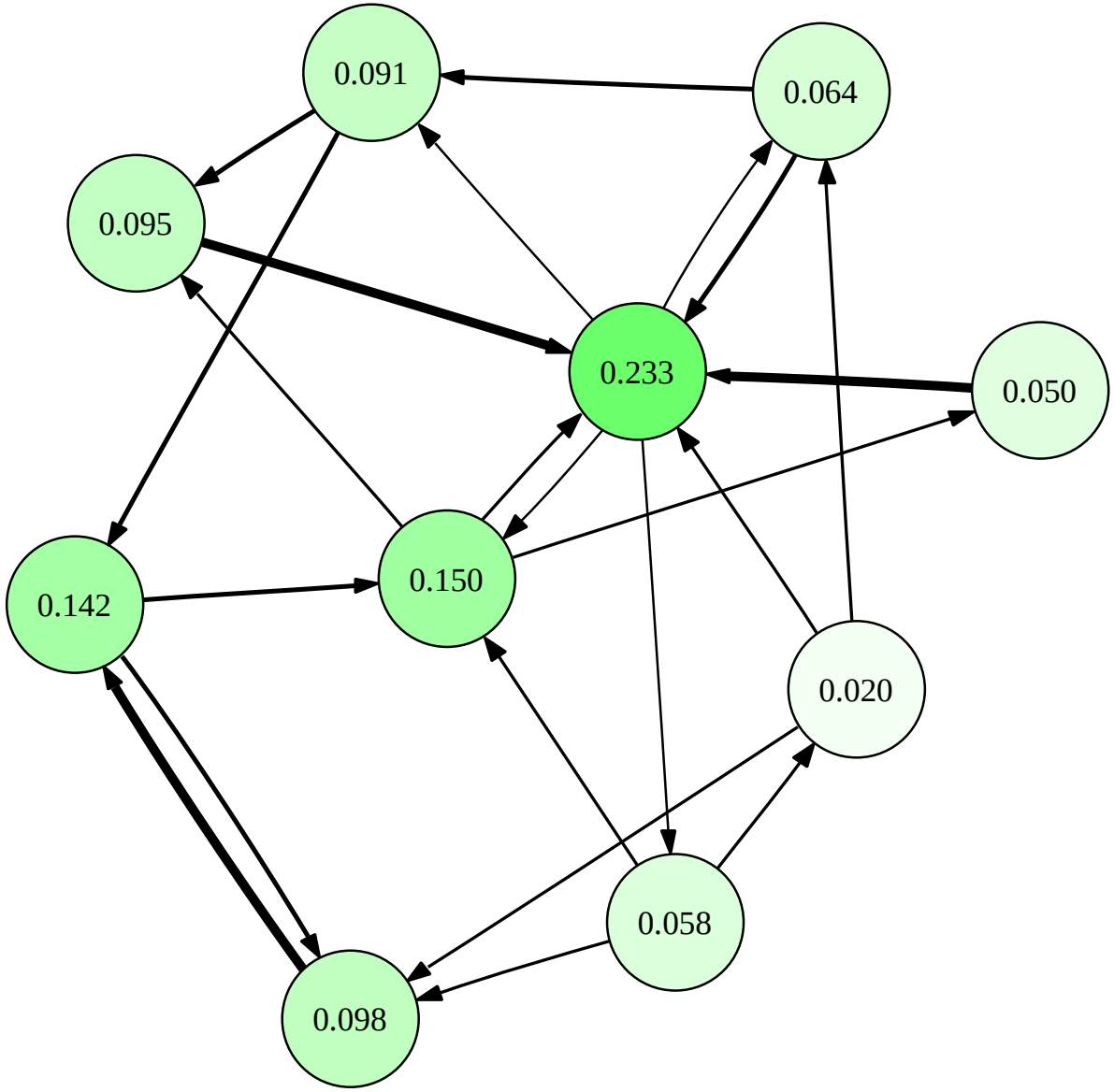
Illustrating PageRank Computation



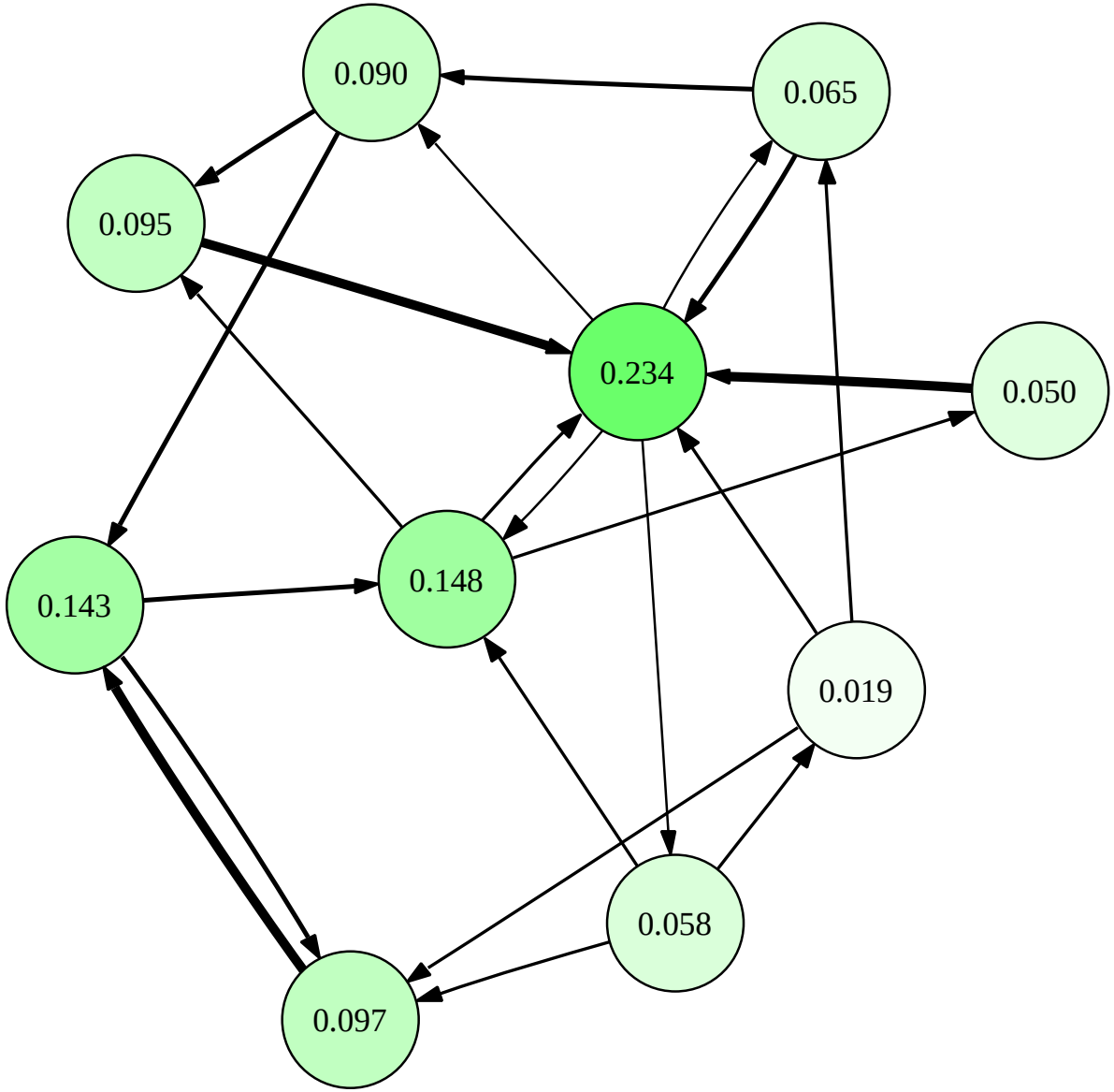
Illustrating PageRank Computation



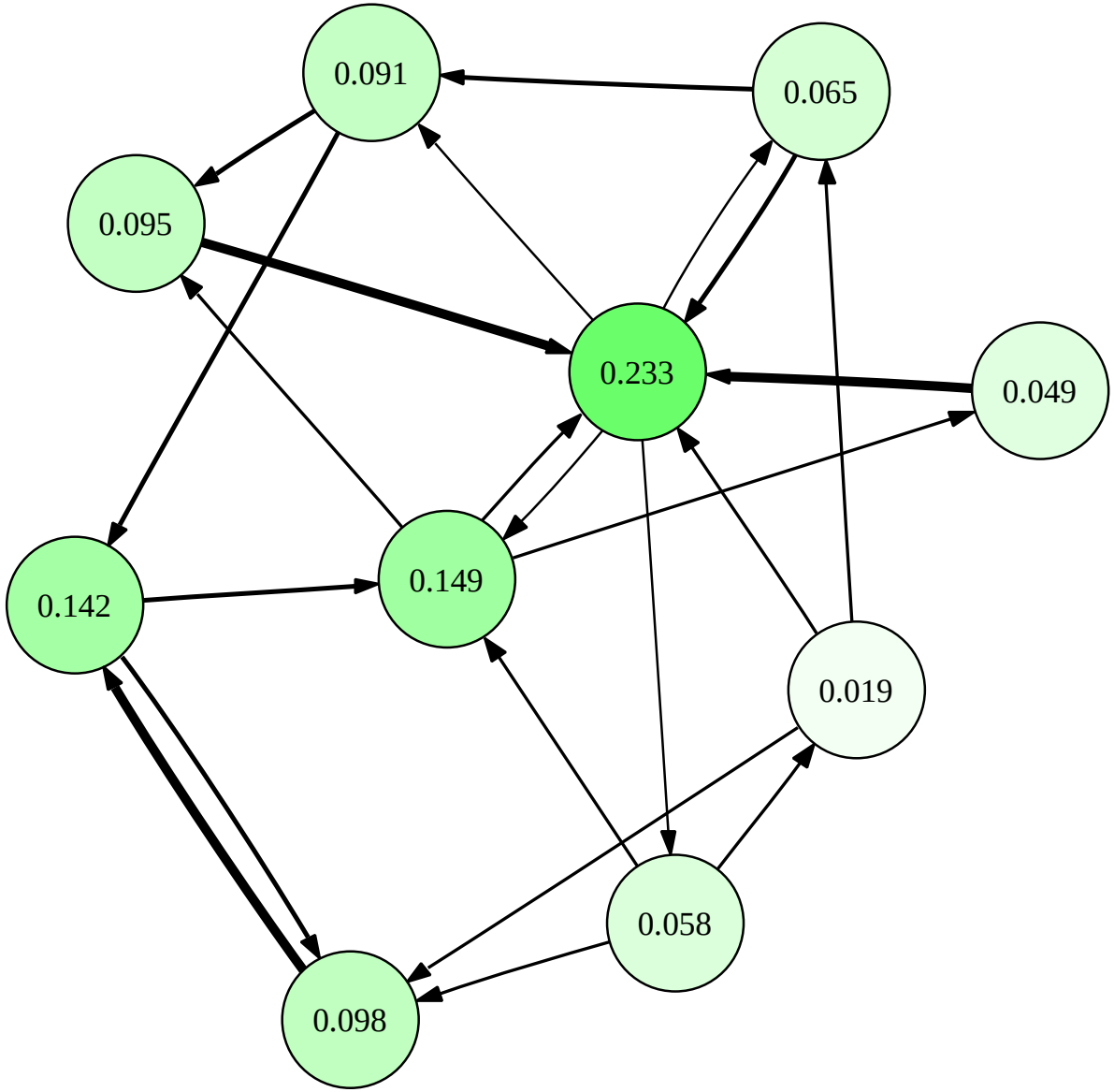
Illustrating PageRank Computation



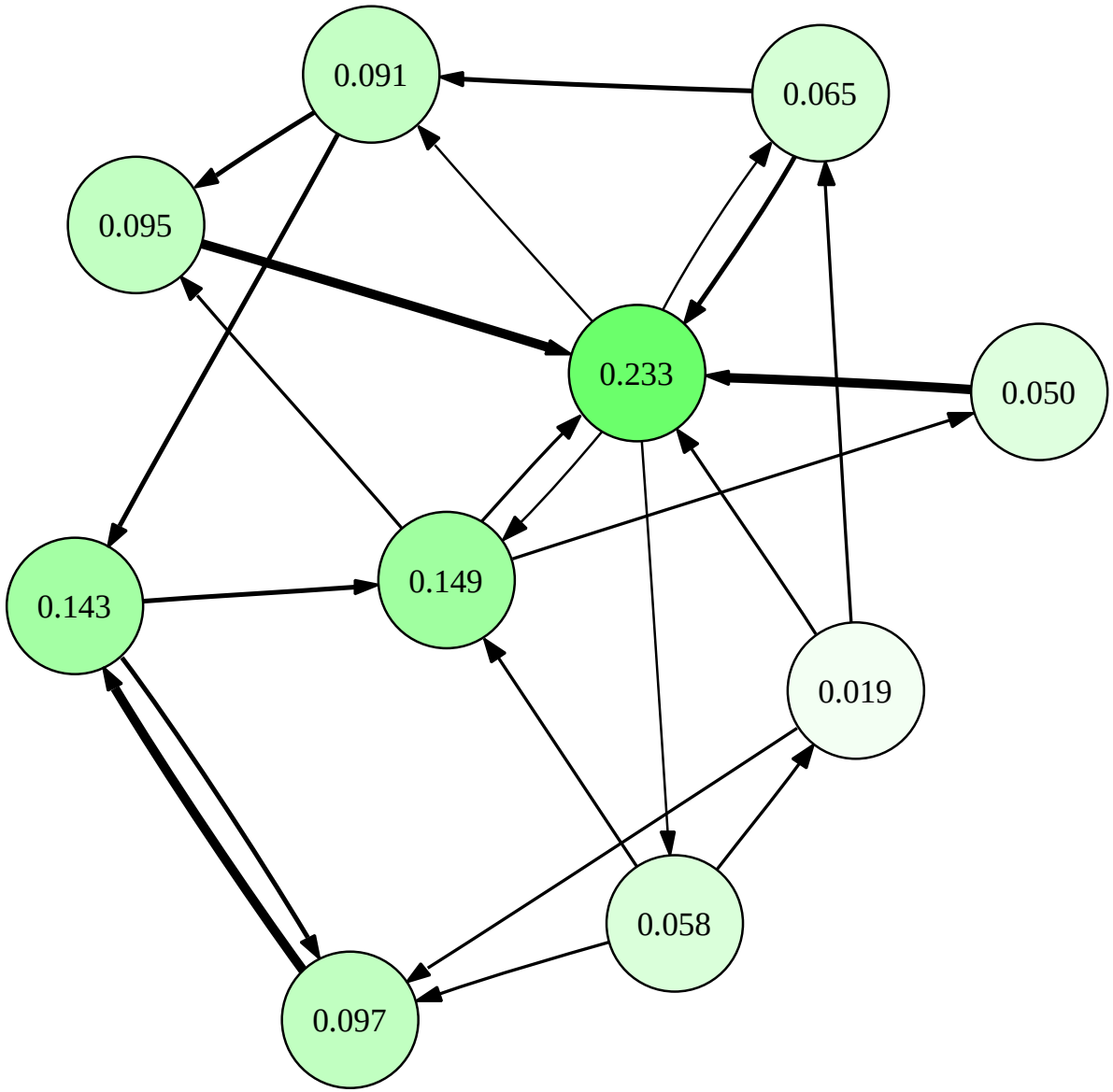
Illustrating PageRank Computation



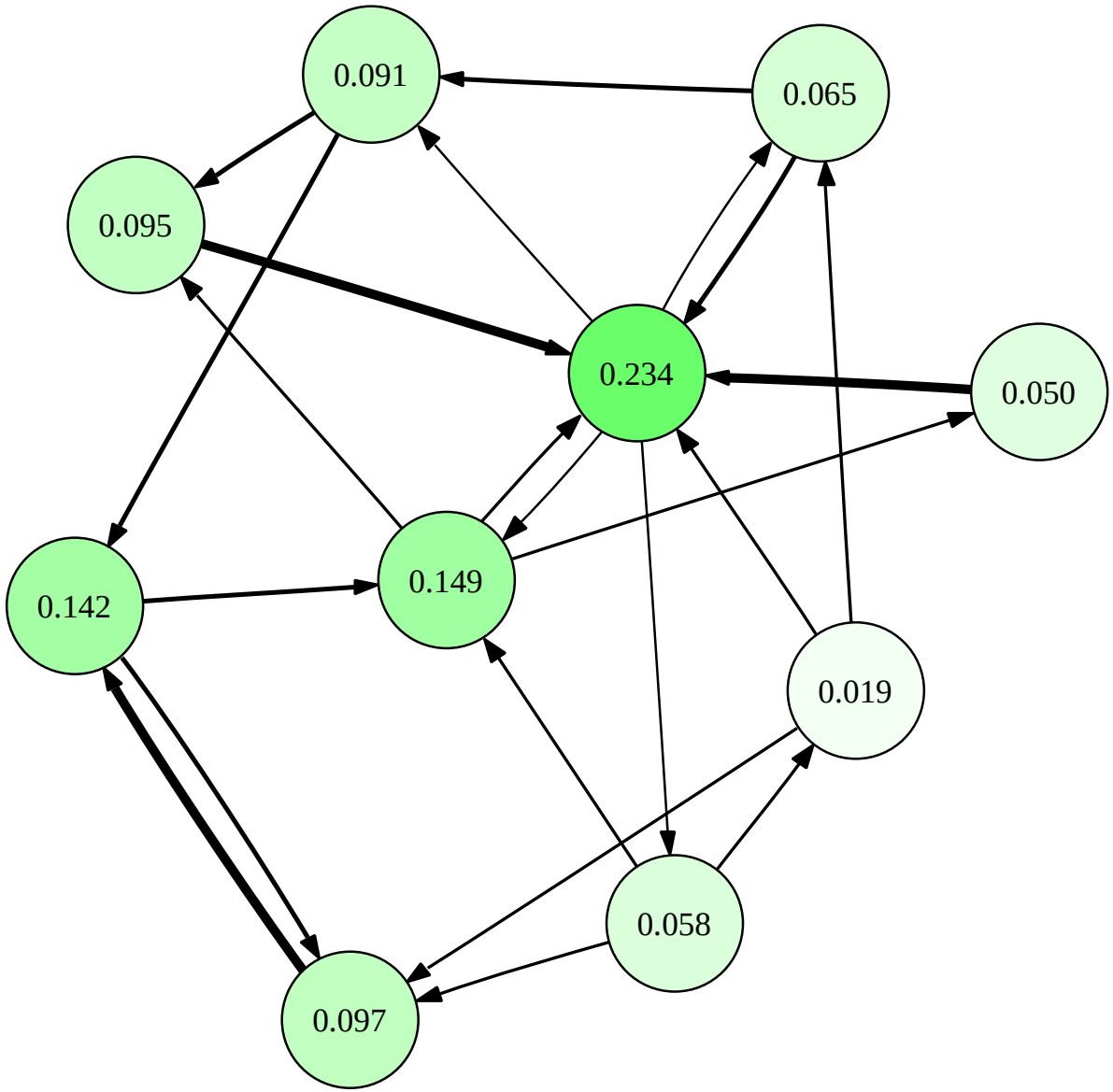
Illustrating PageRank Computation



Illustrating PageRank Computation



Illustrating PageRank Computation



PageRank with Damping

May not always converge, or convergence may not be unique.
To fix this, the random surfer can at each step randomly jump to any page of the Web with some probability d ($1 - d$: damping factor).

$$\mathrm{pr}(i) = \left(\lim_{k \to +\infty} ((1-d)G^T + dU)^k v \right)_i$$

where U is the matrix with all $\frac{1}{N}$ values with N the number of vertices.

Using PageRank to Score Query Results

- PageRank: global score, independent of the query
- Can be used to raise the weight of important pages:

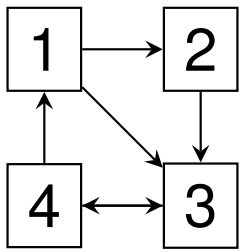
$$\text{weight}(t, d) = \text{tfidf}(t, d) \times \text{pr}(d),$$

- This can be directly incorporated in the index.

Iterative Computation of PageRank

1. Compute G (often stored as its adjacency list). Make sure lines sum to 1.
2. Let u be the uniform vector of sum 1, $v = u$, w the *zero* vector.
3. While v is different enough from w:
   - Set $w = v$.
   - Set $v = (1 - d)G^T v + du$.

Exercise



Run the first iteration of the PageRank computation.

Another take on PageRank: OPIC [Abiteboul et al., 2003]

- Online Page Importance Computation
- Allows computing PageRank without storing the Web graph
- Remember for each page its cash and its history
- Initially, each page has some initial cash
- Each time a page is accessed, distribute uniformly its cash to outgoing links
- PageRank is approximated as its cashflow: history divided by number of visits
- Converge whatever the strategy for crawling the Web is (as soon as each page is accessed repeatedly)
- Good strategy: crawl page with largest cash first

HITS (Kleinberg, [Kleinberg, 1999])

Idea

Two kinds of important pages: hubs and authorities. Hubs are pages that point to good authorities, whereas authorities are pages that are pointed to by good hubs.

G' transition matrix (with 0 and 1 values) of a subgraph of the Web. We use the following iterative process (starting with a and h vectors of norm 1):

$$\begin{cases} a := \frac{1}{\|G'^T h\|} G'^T h \\ h := \frac{1}{\|G' a\|} G' a \end{cases}$$

Converges under some technical assumptions to authority and hub scores.

Using HITS to Order Web Query Results

1. Retrieve the set D of Web pages matching a keyword query.
2. Retrieve the set D^* of Web pages obtained from D by adding all linked pages, as well as all pages linking to pages of D.
3. Build from D^* the corresponding subgraph G' of the Web graph.
4. Compute iteratively hubs and authority scores.
5. Sort documents from D by authority scores.

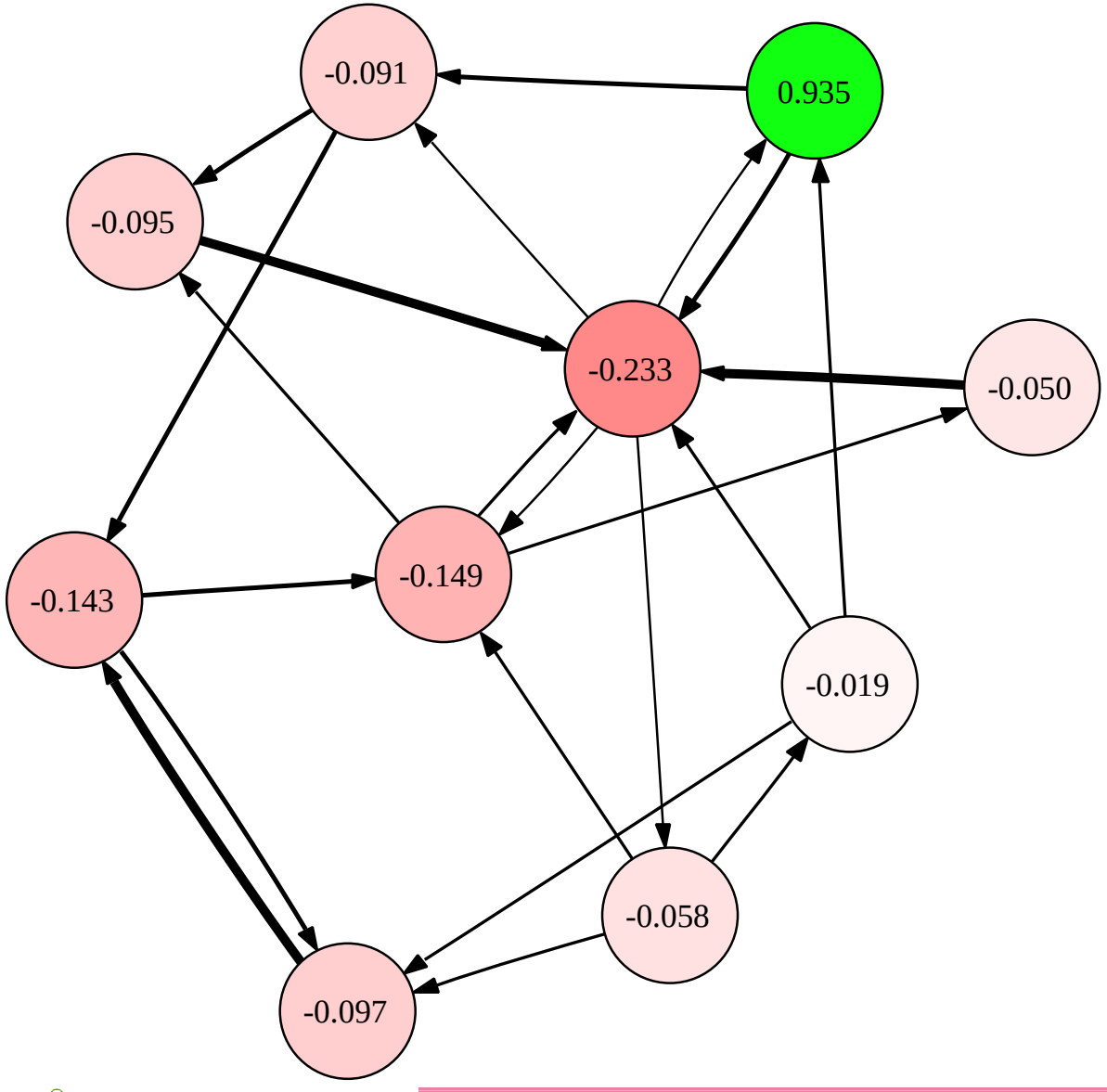
Less efficient than PageRank, because local scores.

Biasing PageRank: Green Measures

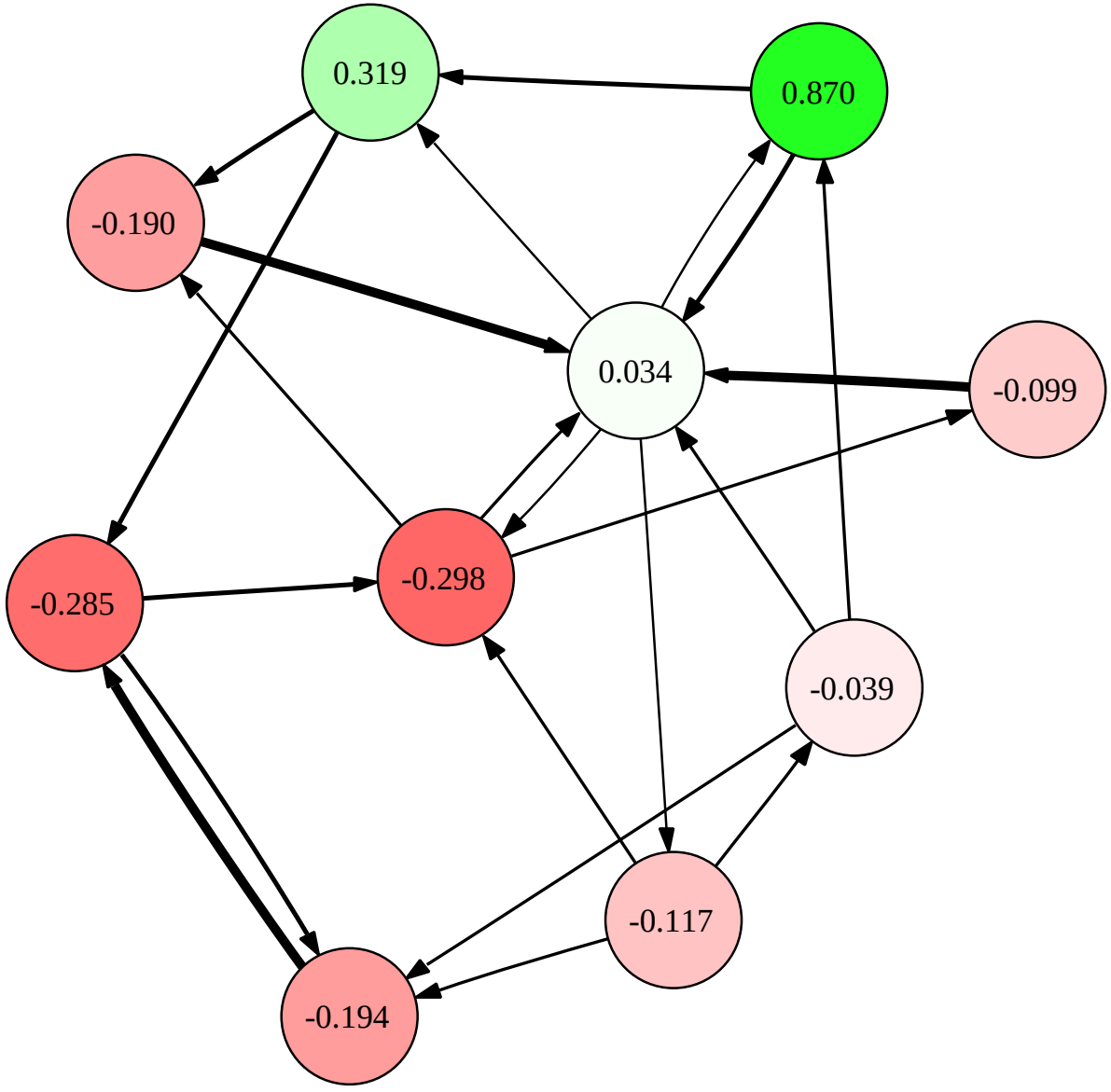
Equivalent definitions of Green measure centered at node i

- PageRank with source at i: standard PageRank computation while, at each iteration, adding 1 to the measure of i, and subtracting $pr(j)$ to every node j.
- Time spent at a node knowing the initial node is i.

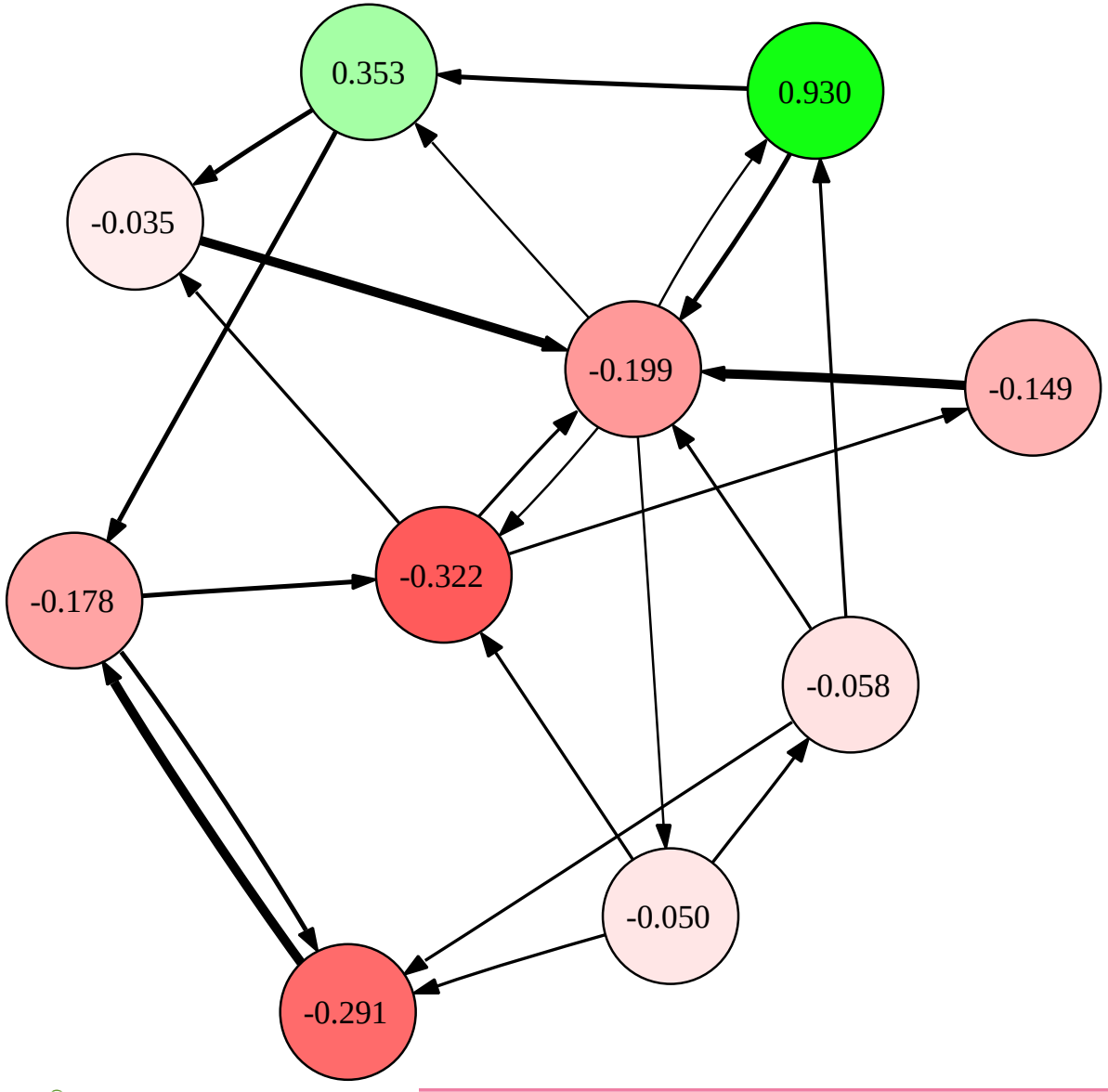
Illustrating Green Measures [Ollivier and Senellart, 2007]



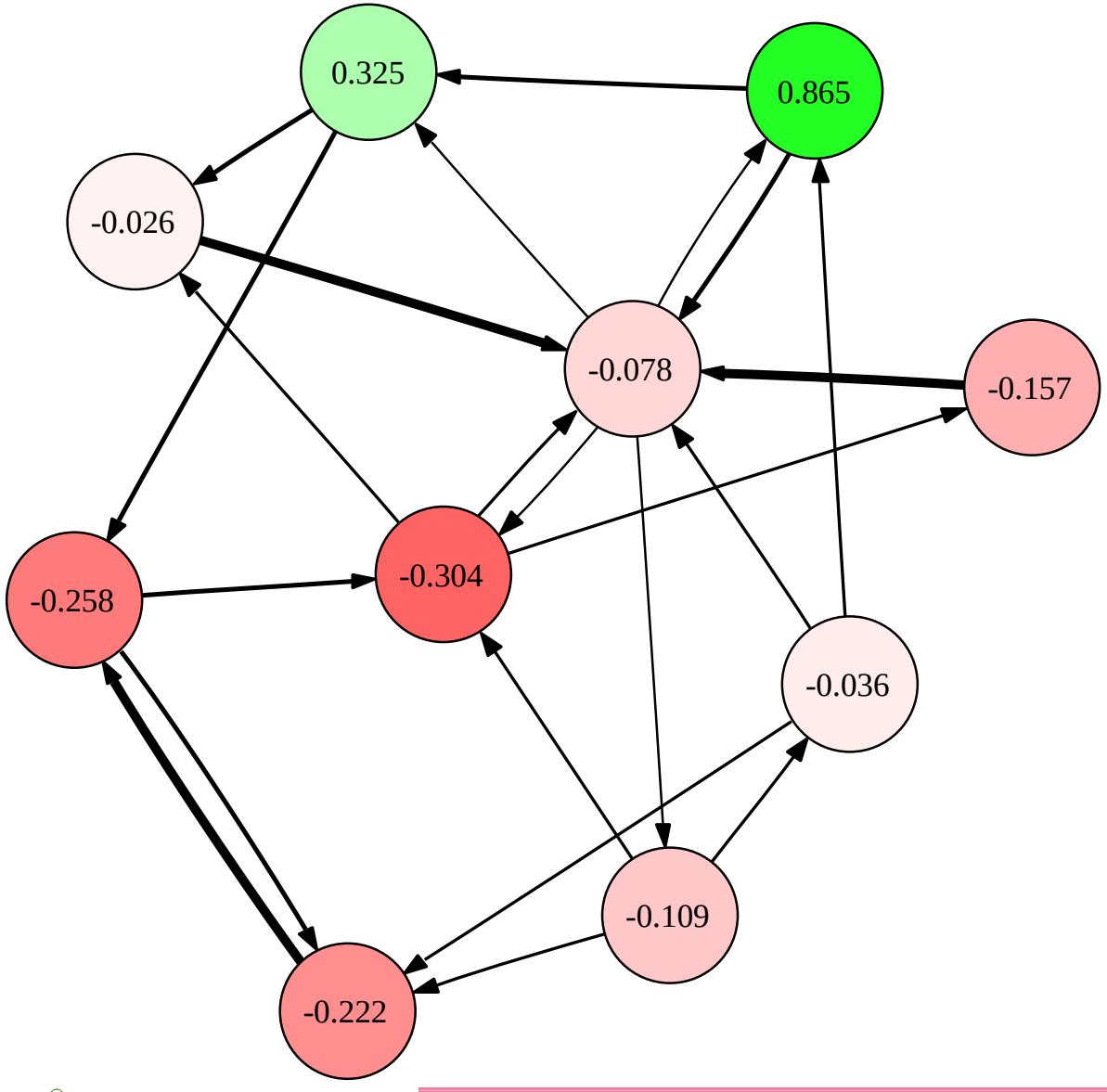
Illustrating Green Measures [Ollivier and Senellart, 2007]



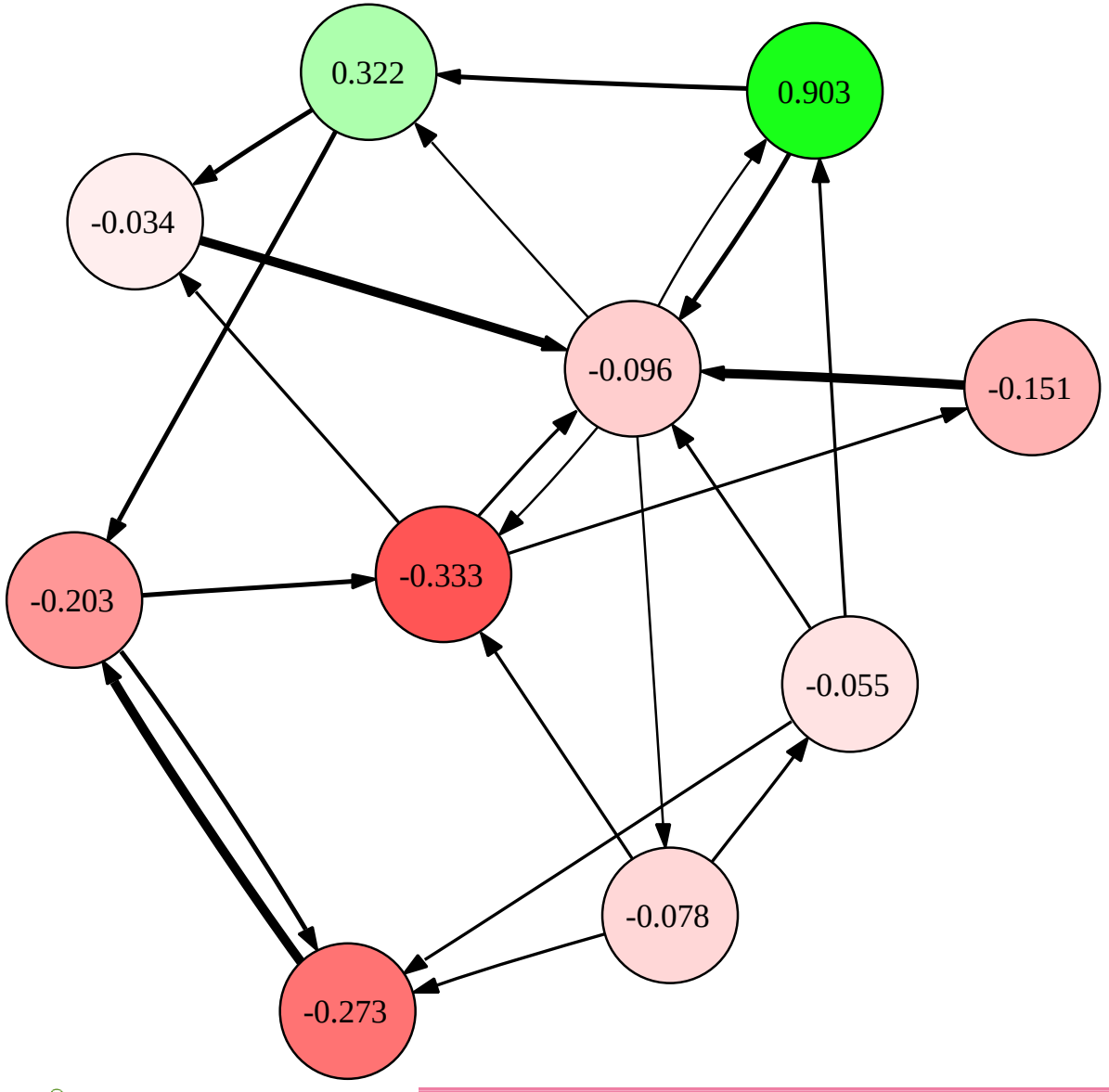
Illustrating Green Measures [Ollivier and Senellart, 2007]



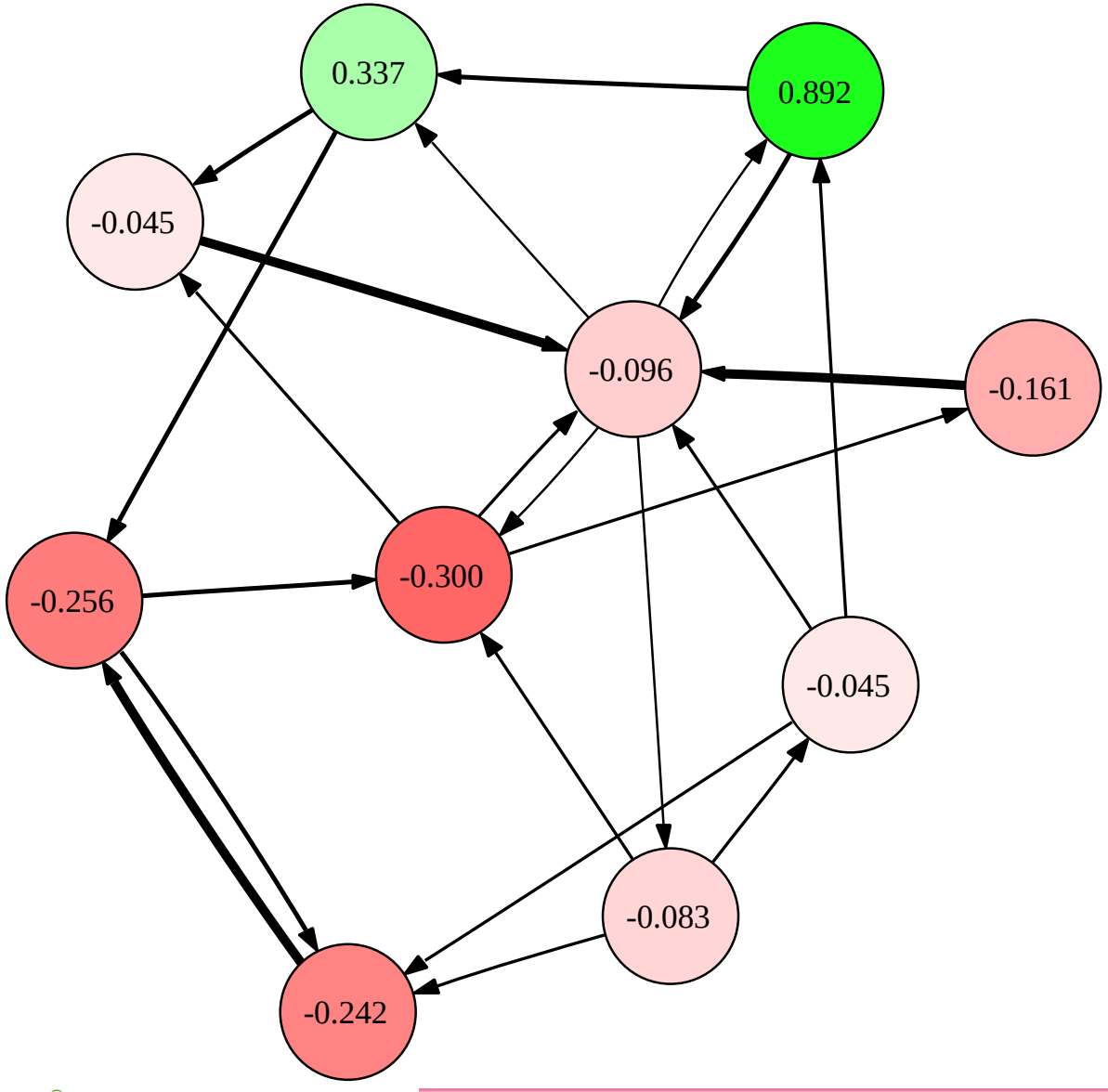
Illustrating Green Measures [Ollivier and Senellart, 2007]



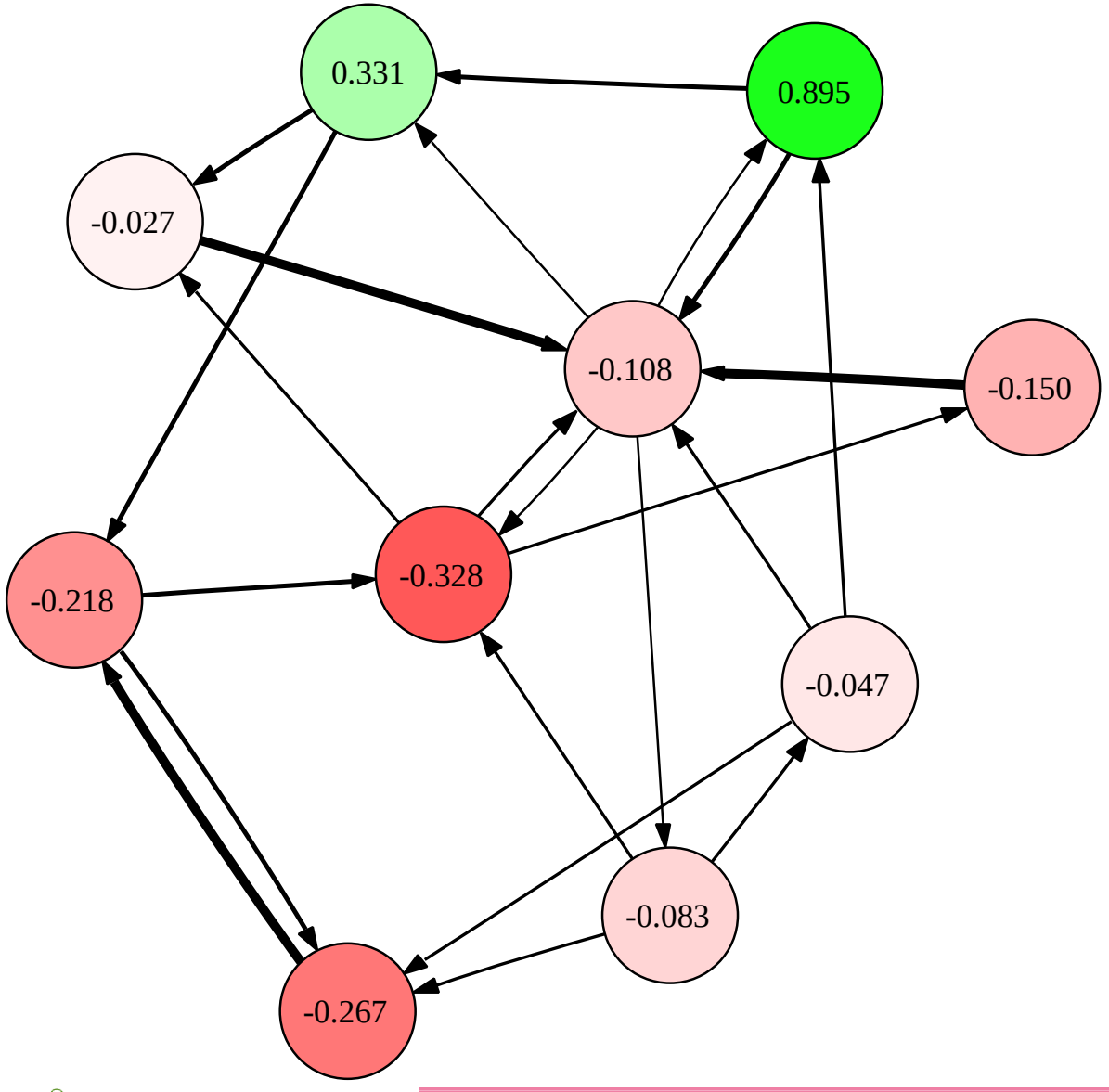
Illustrating Green Measures [Ollivier and Senellart, 2007]



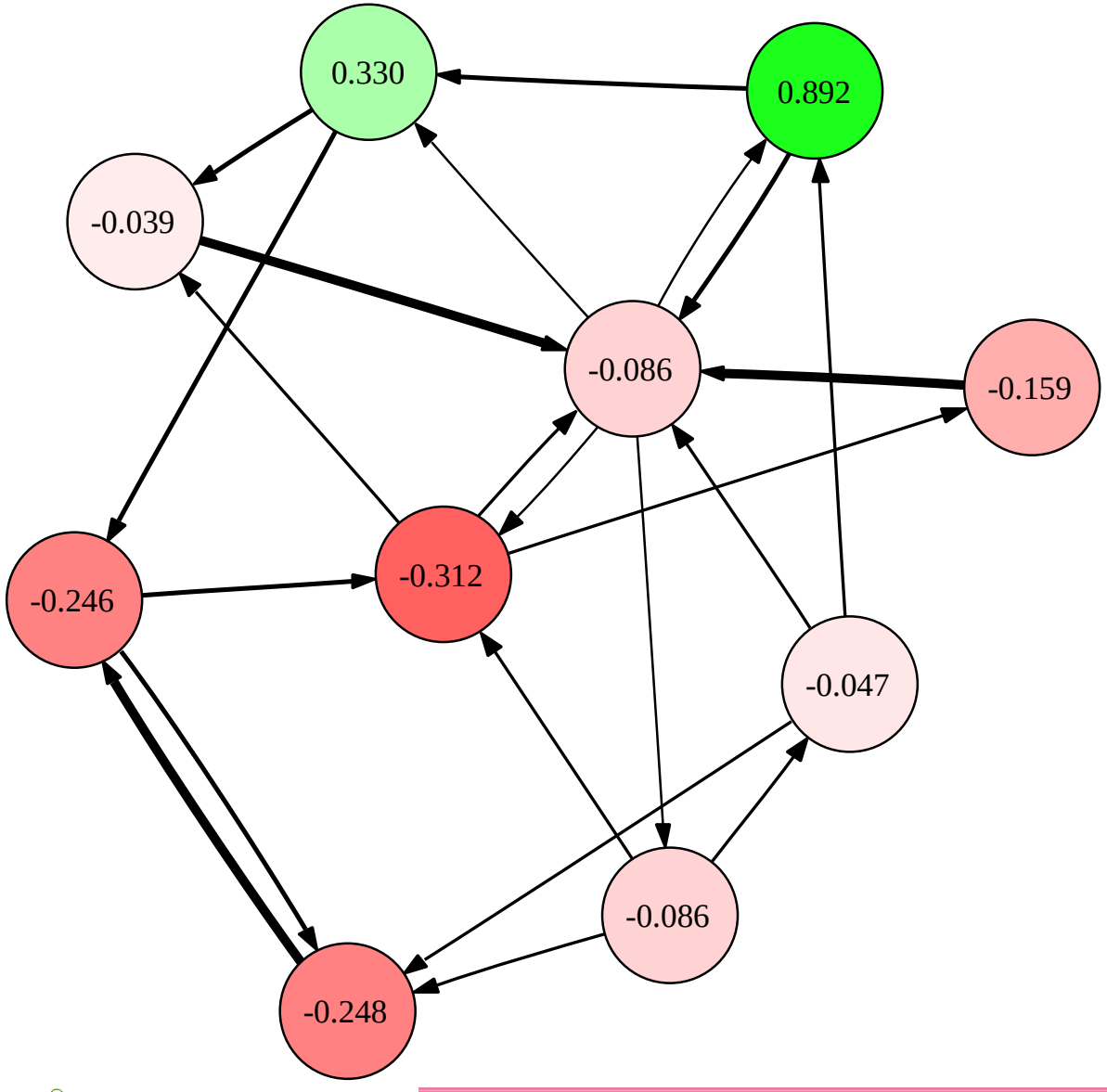
Illustrating Green Measures [Ollivier and Senellart, 2007]



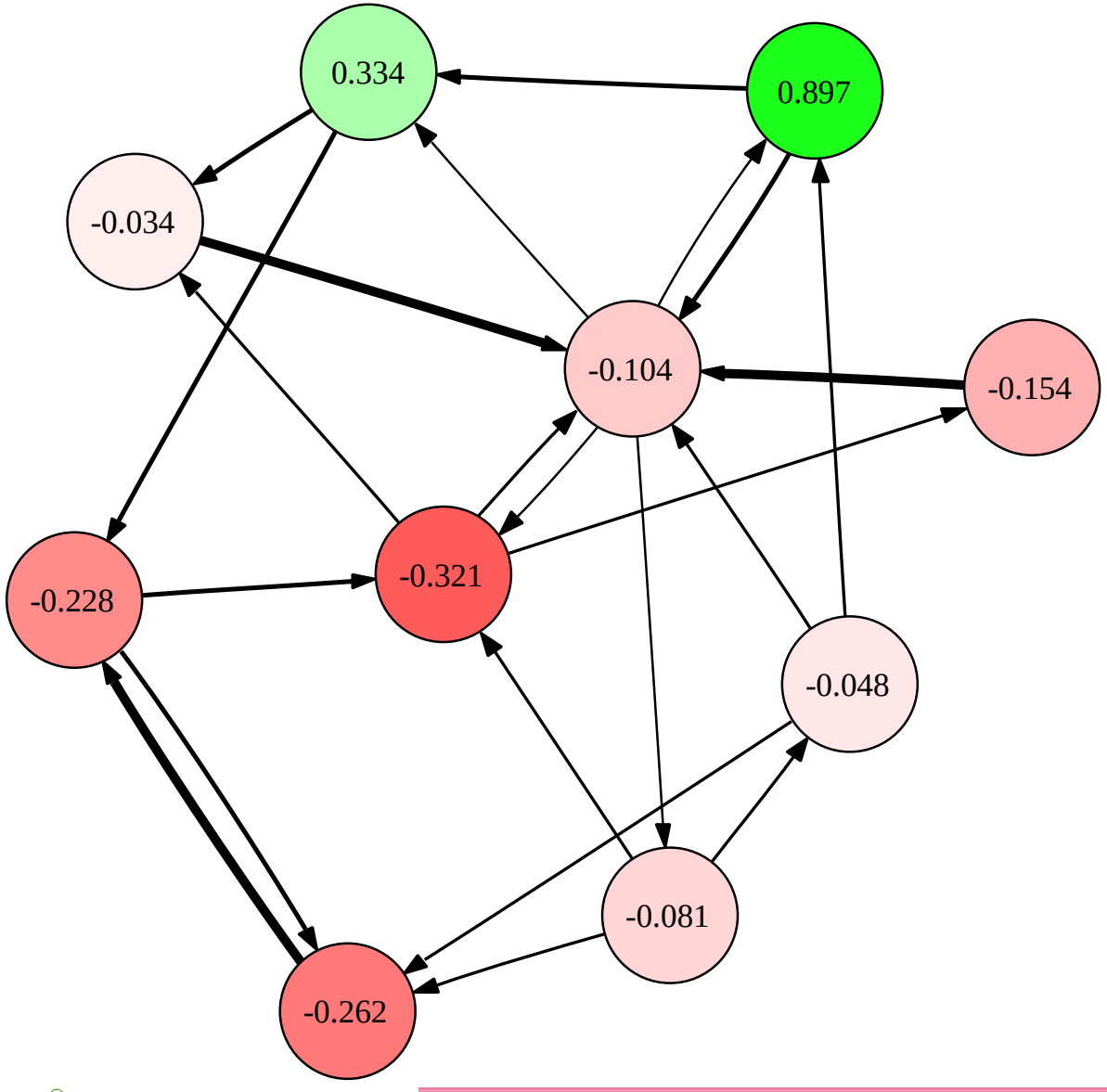
Illustrating Green Measures [Ollivier and Senellart, 2007]



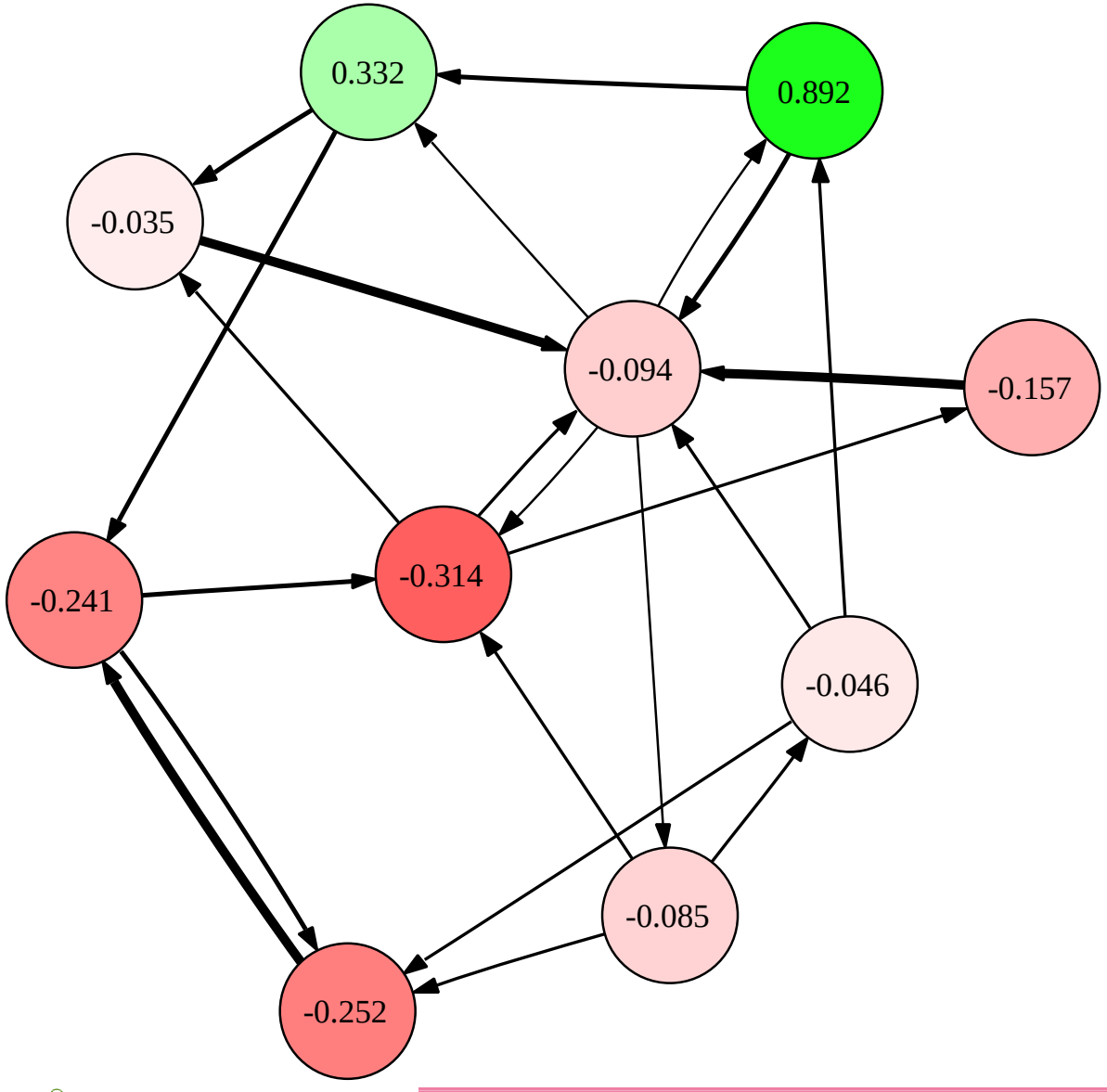
Illustrating Green Measures [Ollivier and Senellart, 2007]



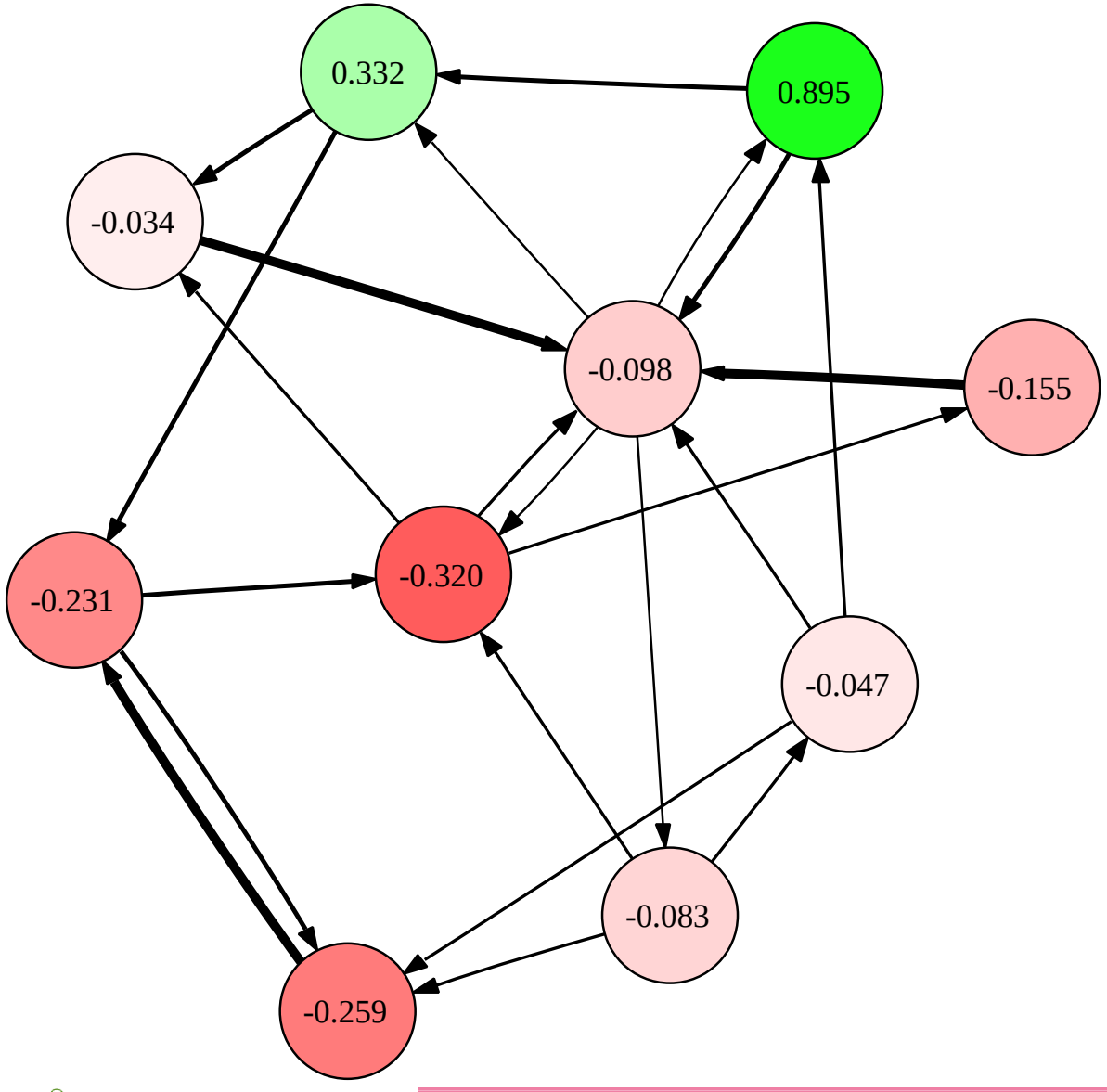
Illustrating Green Measures [Ollivier and Senellart, 2007]



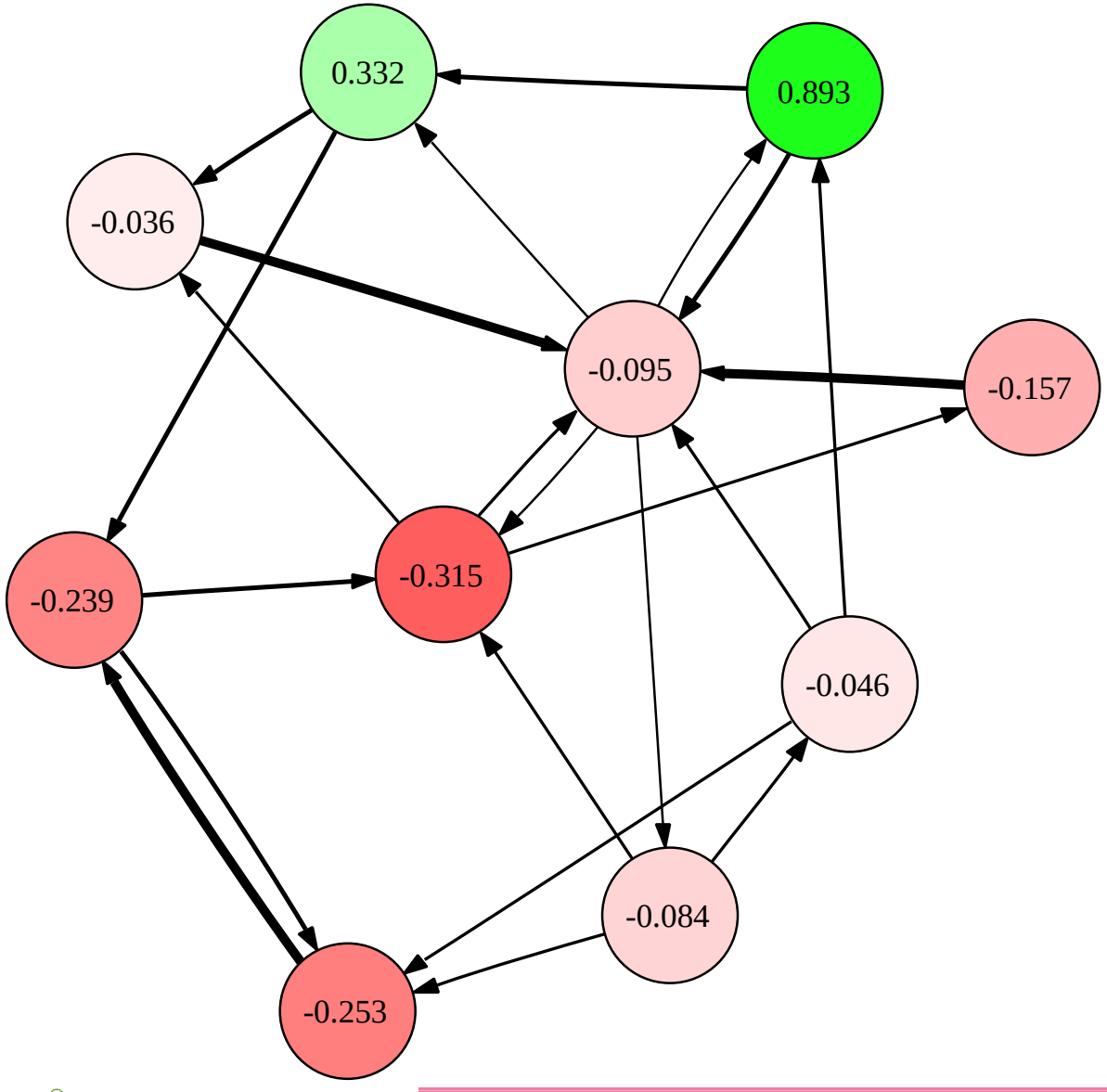
Illustrating Green Measures [Ollivier and Senellart, 2007]



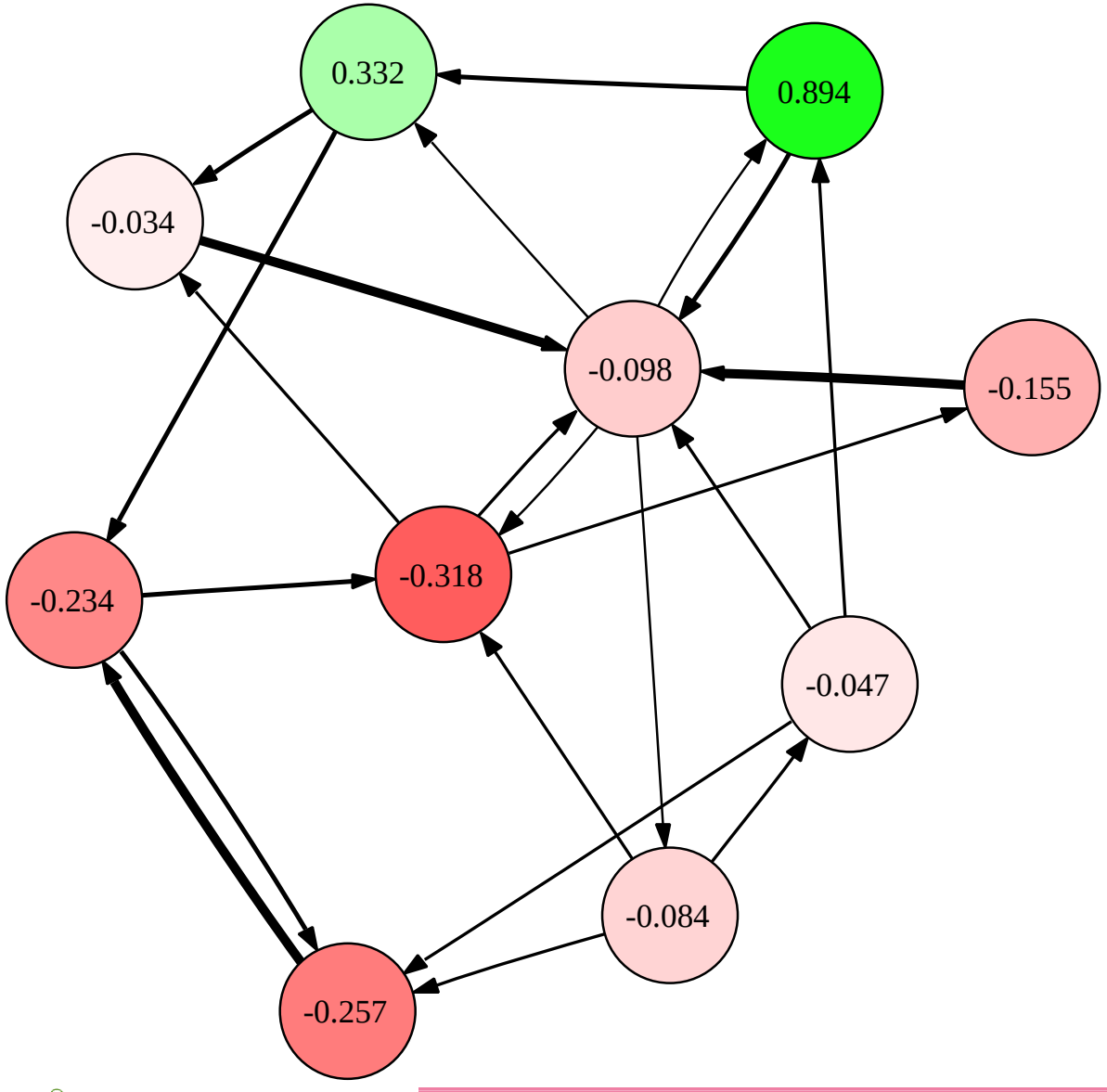
Illustrating Green Measures [Ollivier and Senellart, 2007]



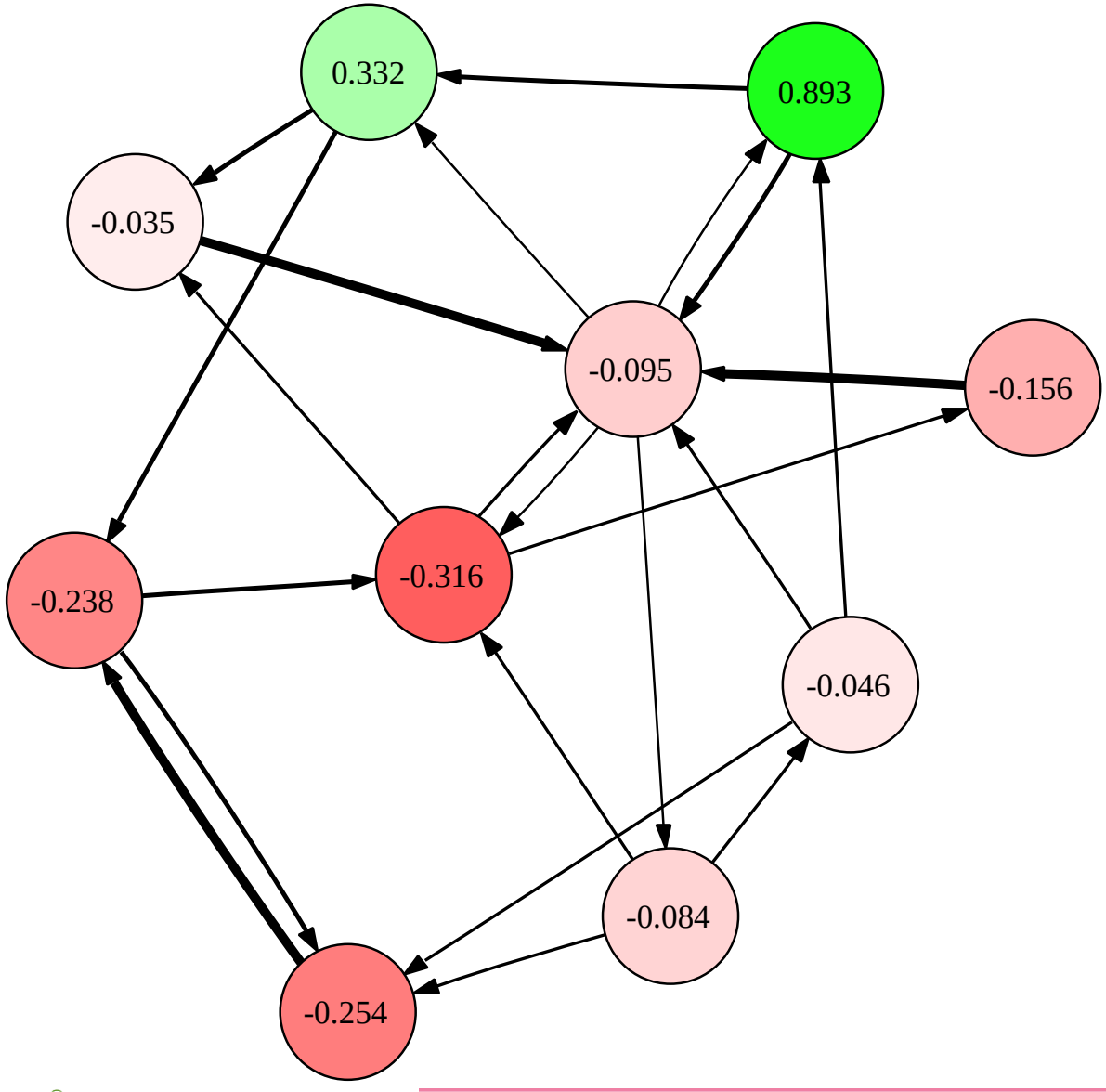
Illustrating Green Measures [Ollivier and Senellart, 2007]



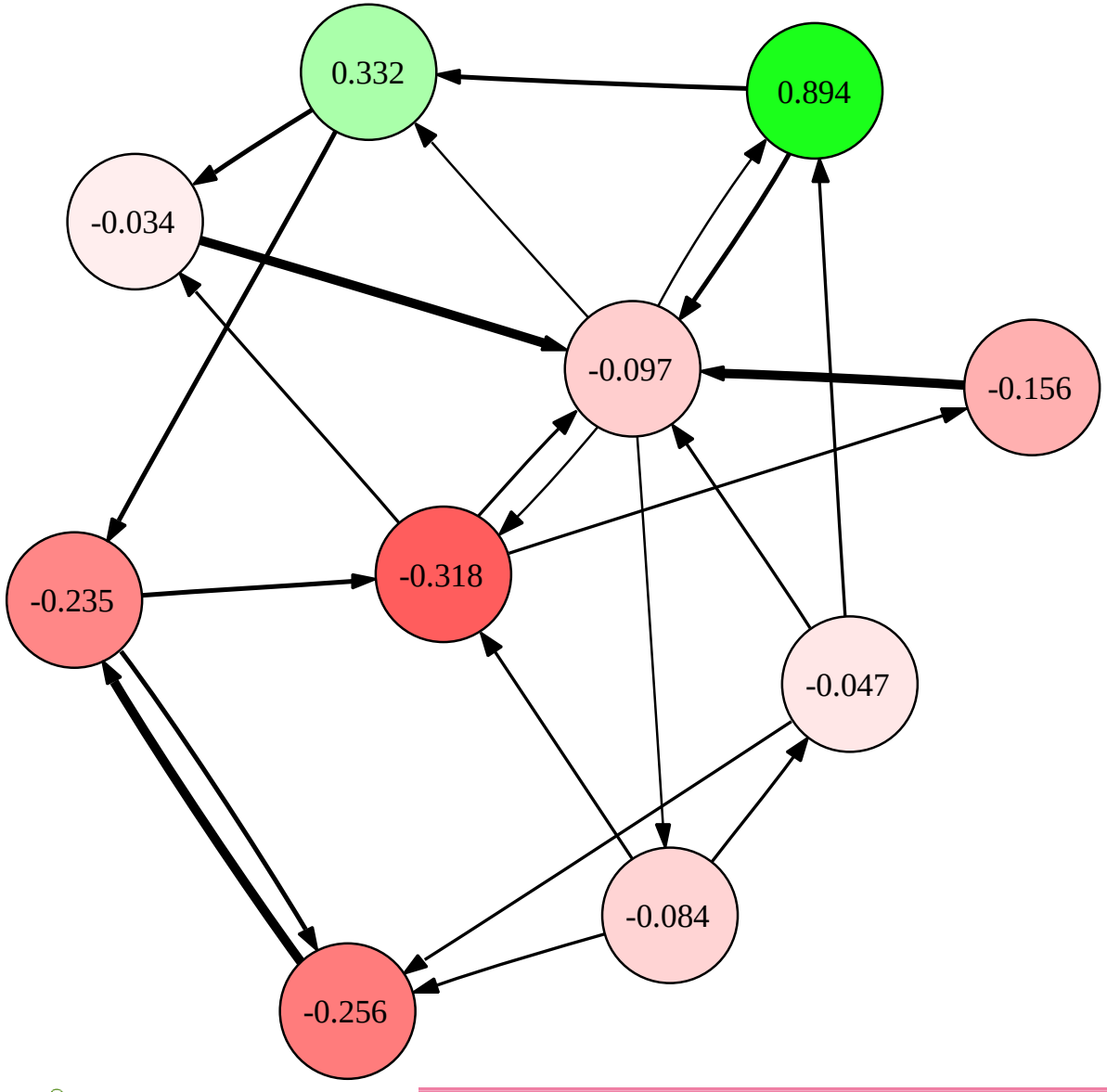
Illustrating Green Measures [Ollivier and Senellart, 2007]



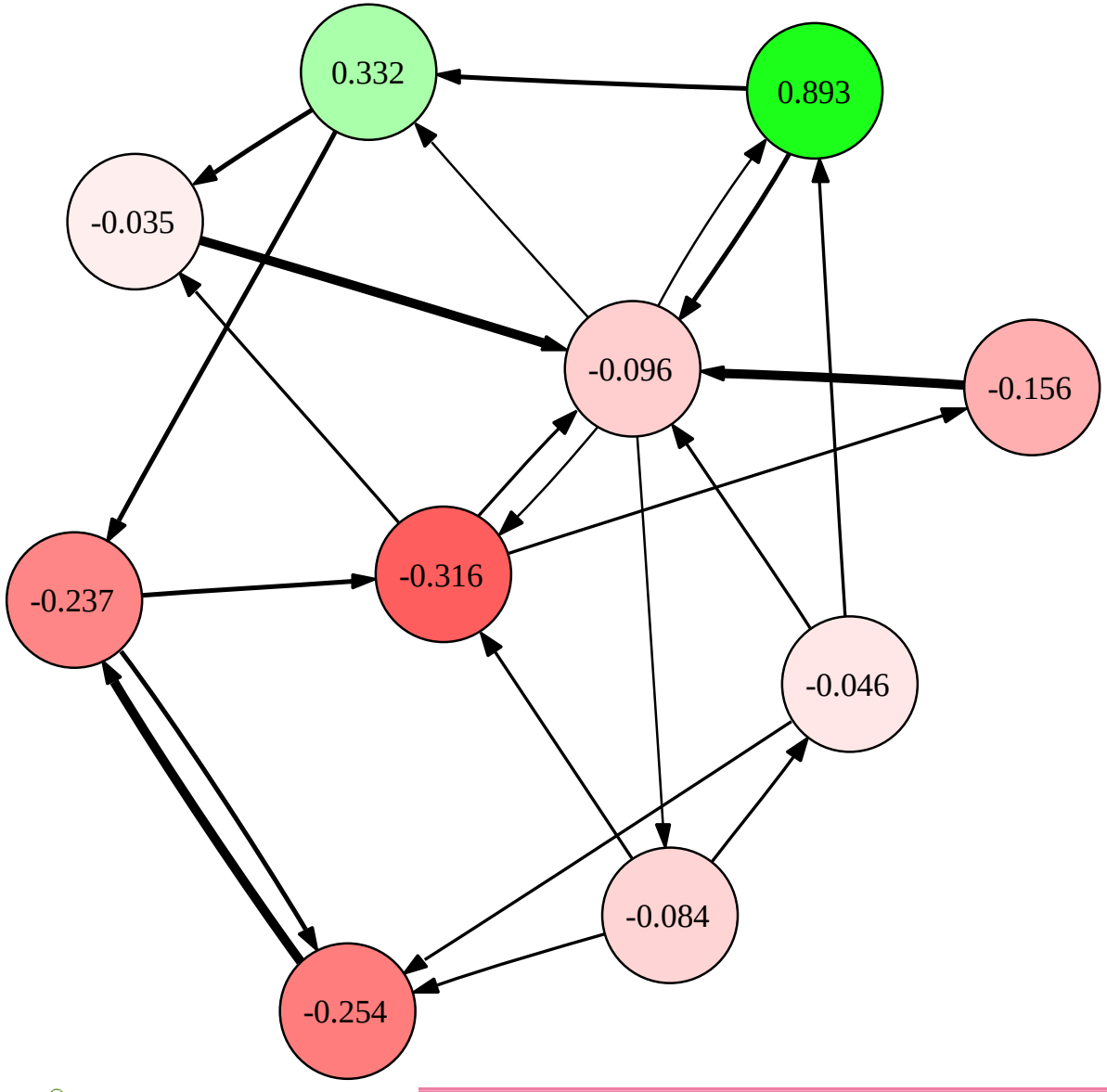
Illustrating Green Measures [Ollivier and Senellart, 2007]



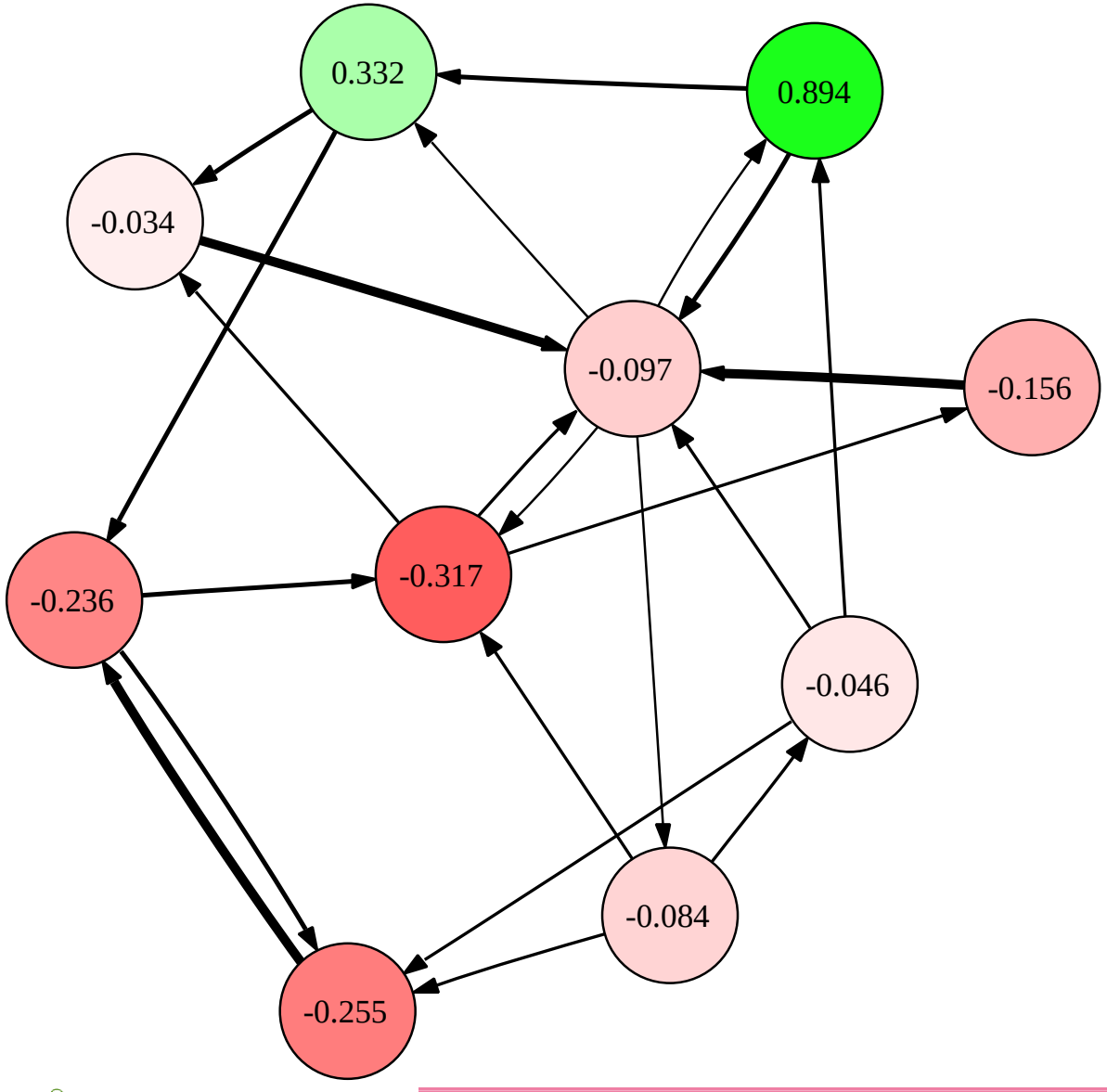
Illustrating Green Measures [Ollivier and Senellart, 2007]



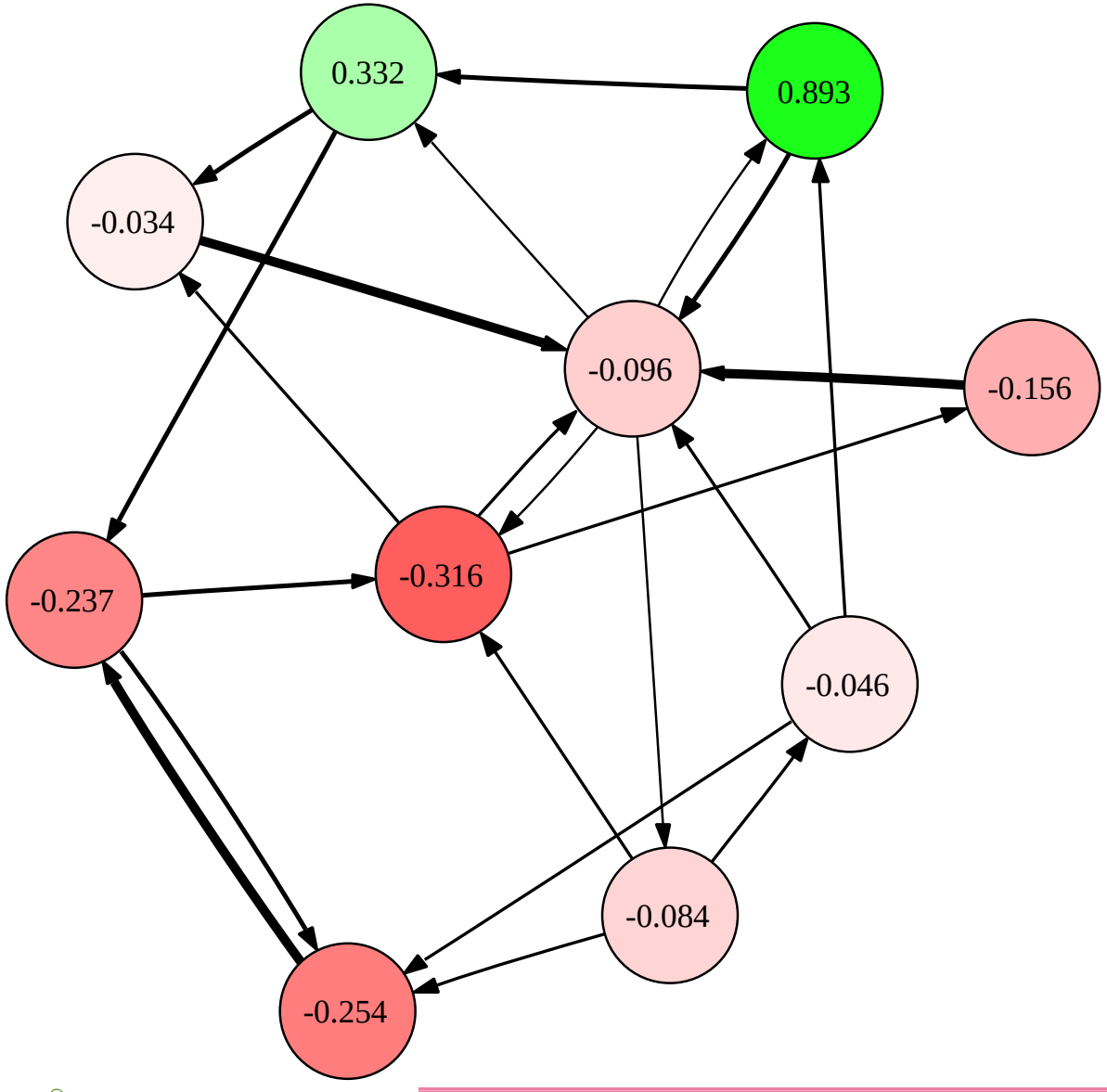
Illustrating Green Measures [Ollivier and Senellart, 2007]



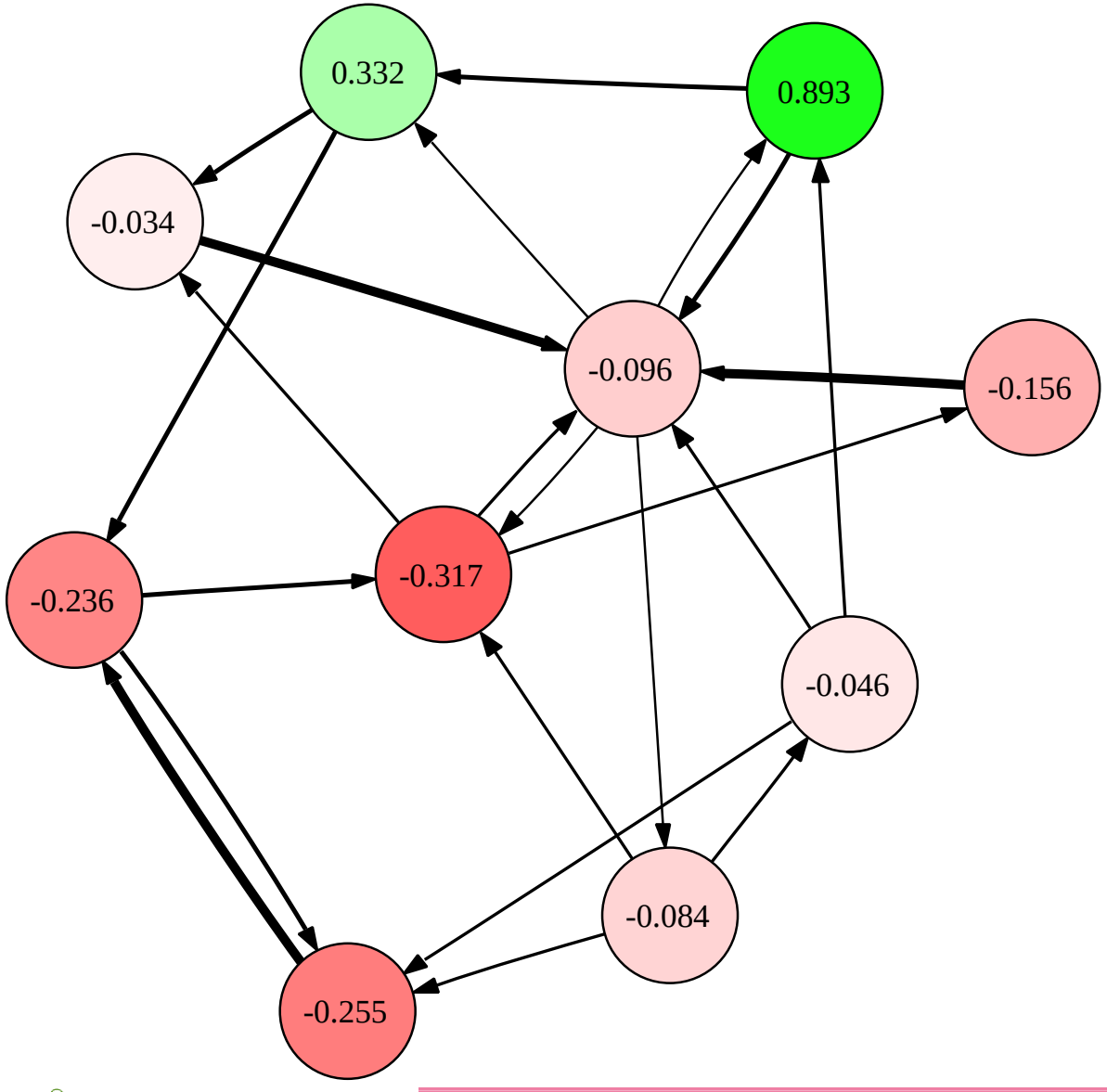
Illustrating Green Measures [Ollivier and Senellart, 2007]



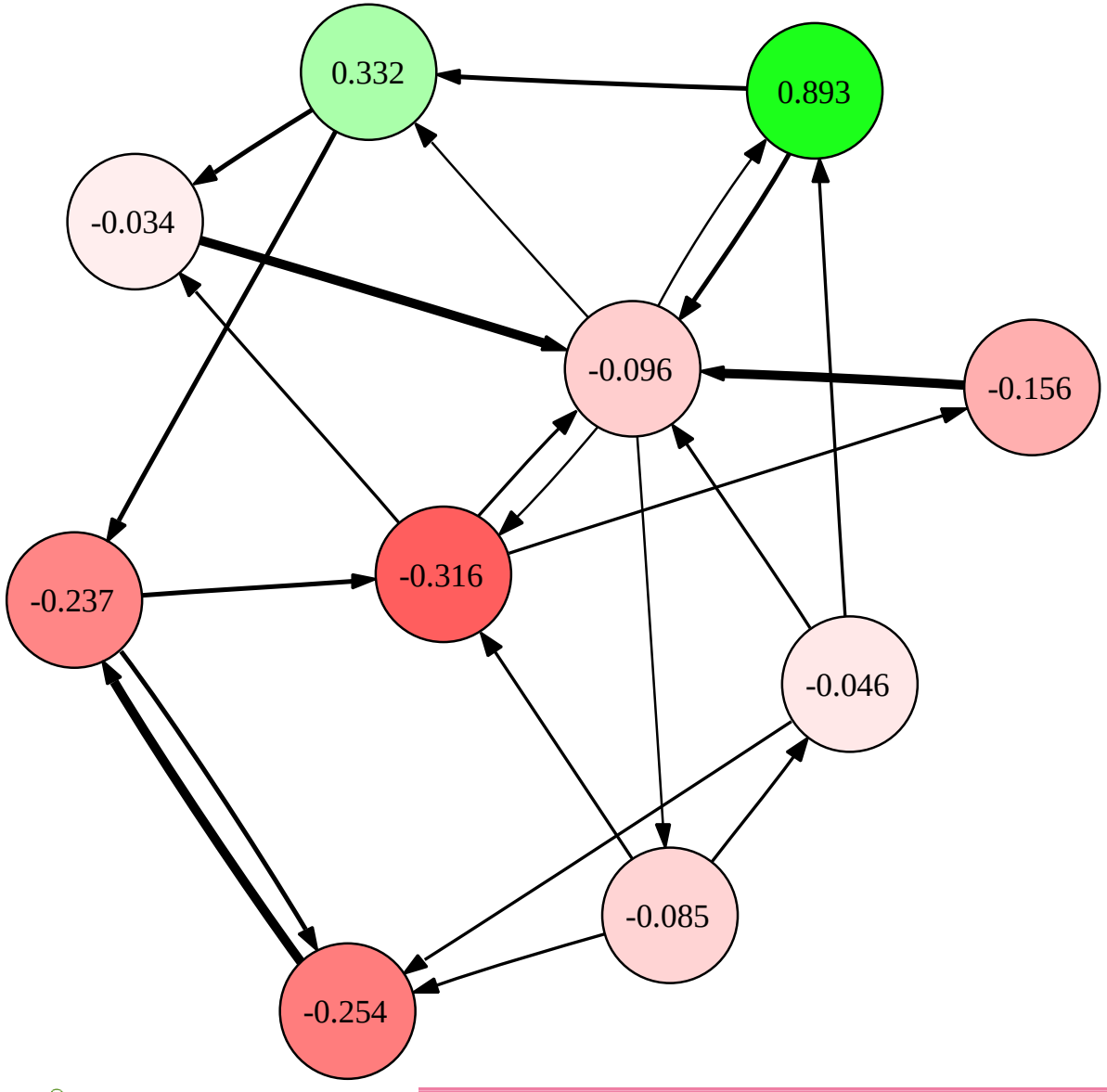
Illustrating Green Measures [Ollivier and Senellart, 2007]



Illustrating Green Measures [Ollivier and Senellart, 2007]



Illustrating Green Measures [Ollivier and Senellart, 2007]



Outline

Basics of Graph Theory

The Web as a graph
- Introduction
- Computing the importance of a page
- Spamdexing
- Discovering communities

Social Networking Sites

Conclusion

Spamdexing

Definition

Fraudulent techniques that are used by unscrupulous webmasters to artificially raise the visibility of their website to users of search engines

Purpose: attracting visitors to websites to make profit.

Unceasing war between spamdexers and search engines

Spamdexing: Lying about the Content

Technique

Put unrelated terms in:

- meta-information (`<meta name="description">`, `<meta name="keywords">`)
- text content hidden to the user with JavaScript, CSS, or HTML presentational elements

Countertechnique

- Ignore meta-information
- Try and detect invisible text

Link Farm Attacks

Technique

Huge number of hosts on the Internet used for the sole purpose of referencing each other, without any content in themselves, to raise the importance of a given website or set of websites.

Countertechnique

- Detection of websites with empty or duplicate content
- Use of heuristics to discover subgraphs that look like link farms

Link Pollution

Technique

Pollute user-editable websites (blogs, wikis) or exploit security bugs to add artificial links to websites, in order to raise its importance.

Countertechnique

`rel="nofollow"` attribute to `<a>` links not validated by a page's owner

Outline

Basics of Graph Theory

The Web as a graph
- Introduction
- Computing the importance of a page
- Spamdexing
- Discovering communities

Social Networking Sites

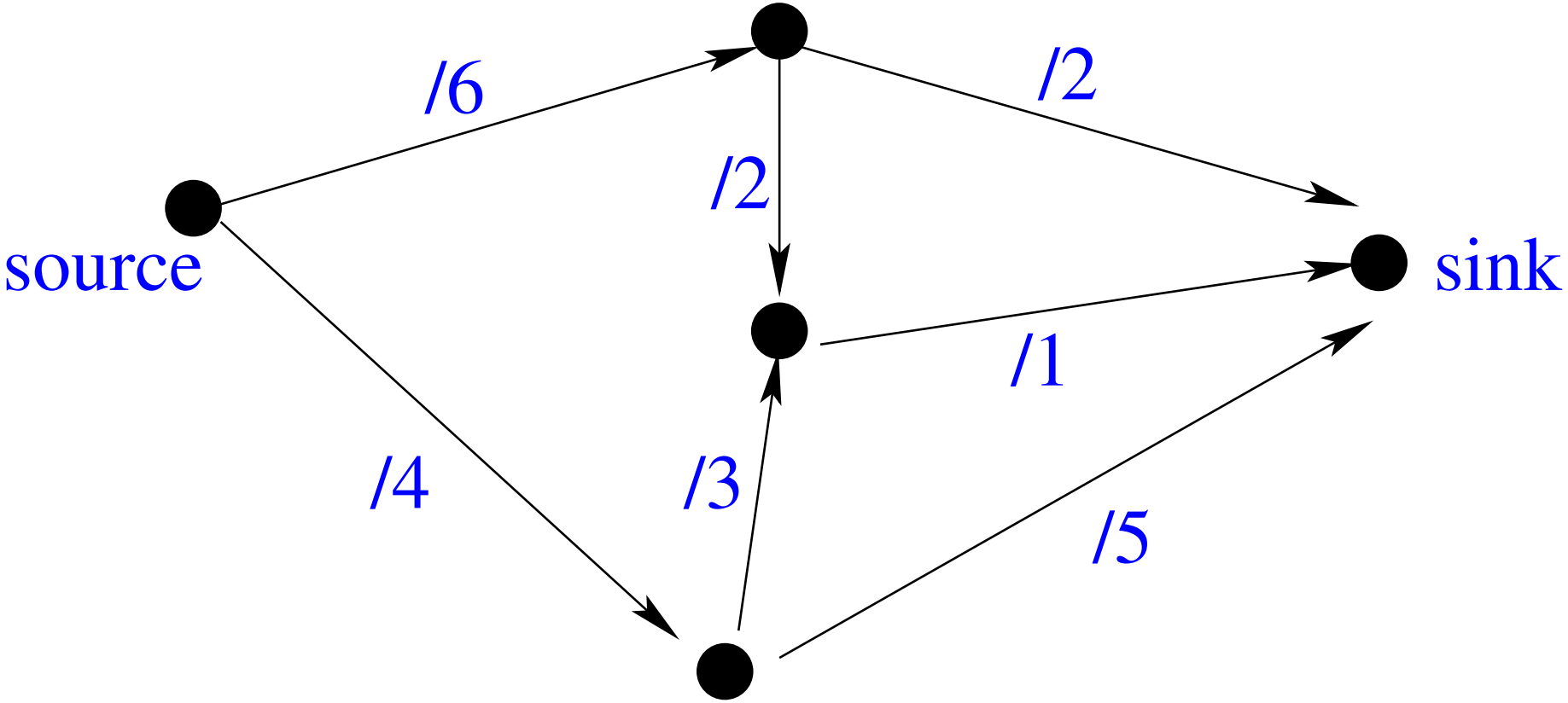
Conclusion

Discovery of communities

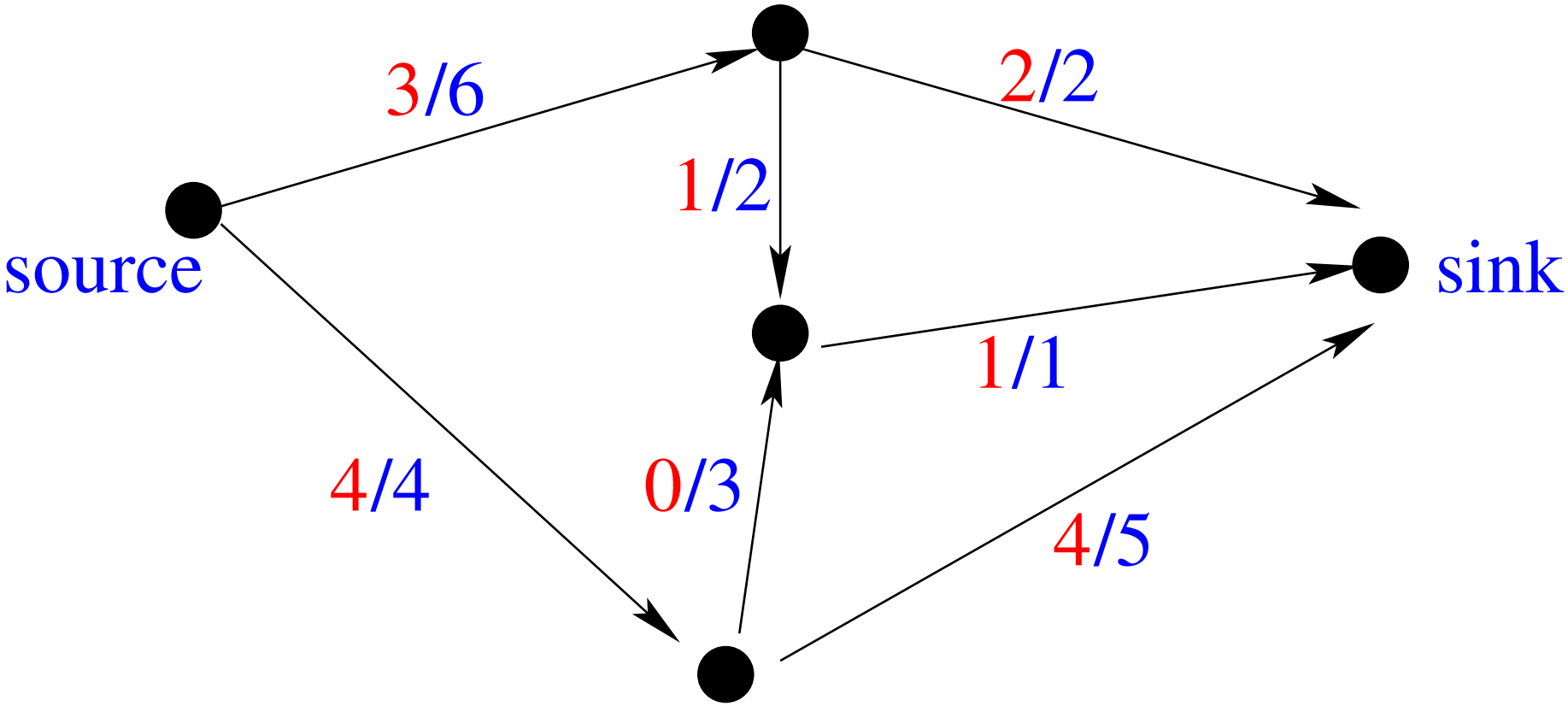
- Identifying communities of Web pages using the graph structure: can be used for clustering the Web, or for finding boundaries of logical Web sites

- Two subproblems:
  1. Given some initial vertex or vertex set, finding the corresponding community
  2. Given the graph as a whole, finding a partition in communities

Maximum Flow / Minimum Cut

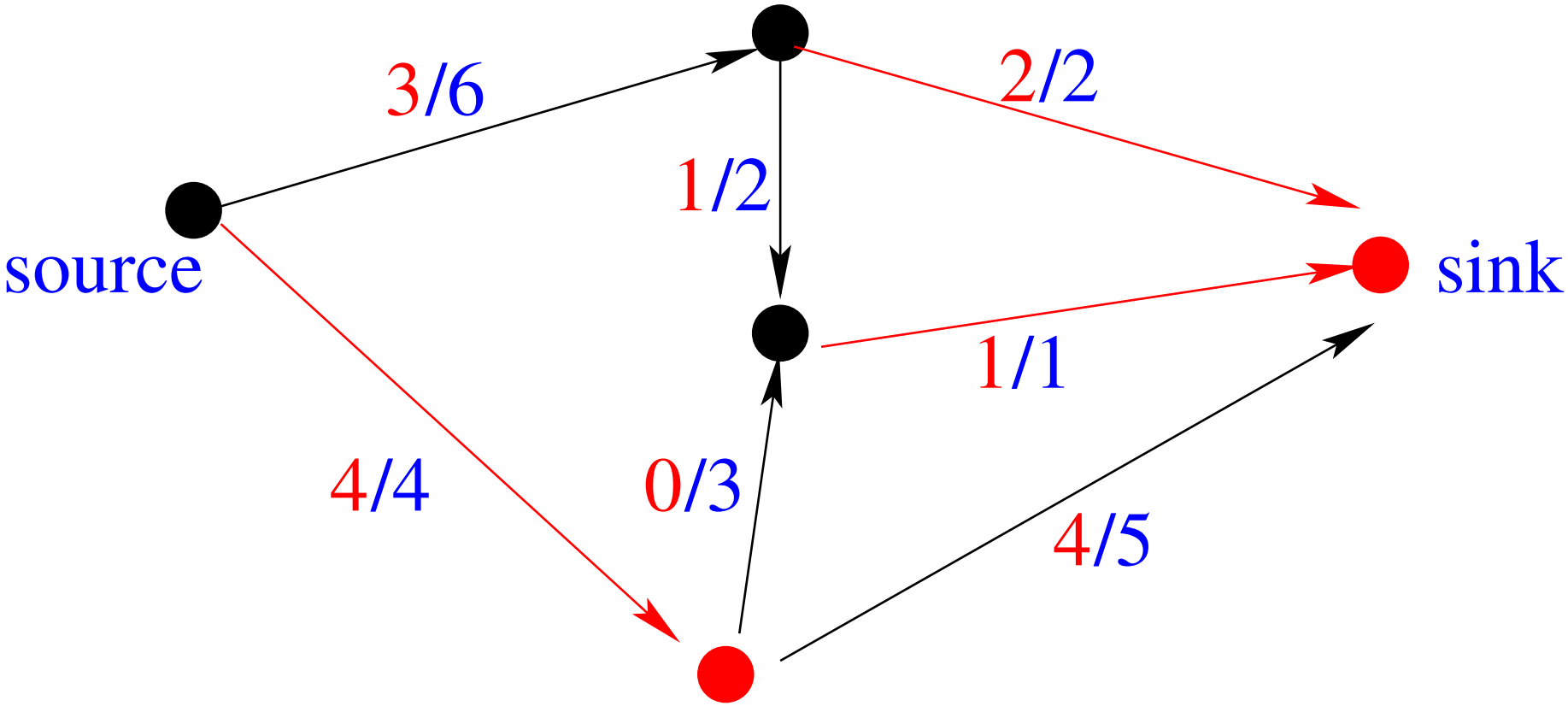


Maximum Flow / Minimum Cut



- Use of a maximum flow computation algorithm [Goldberg and Tarjan, 1988] to separate a seed of users from the remaining of the graph
- Complexity $O(n^2 m)$ (n: vertices, m: edges)

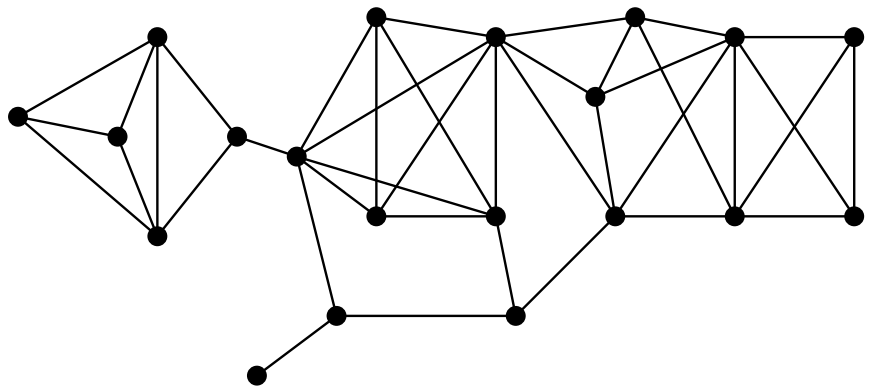
Maximum Flow / Minimum Cut



- Use of a maximum flow computation algorithm [Goldberg and Tarjan, 1988] to separate a seed of users from the remaining of the graph
- Complexity $O(n^2 m)$ (n: vertices, m: edges)

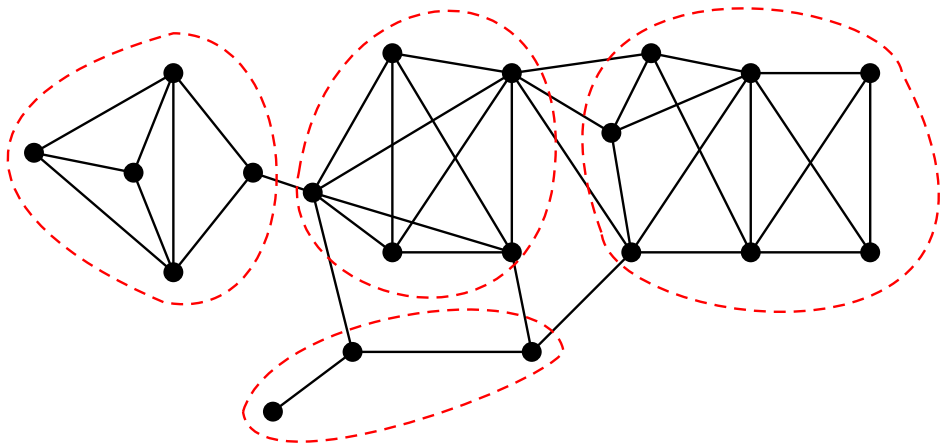
Markov Cluster Algorithm (MCL) [van Dongen, 2000]

- Graph clustering algorithm
- Based as well on maximum flow simulation, in the whole graph
- Iteration of a matrix computation alternating:
  - Expansion (matrix multiplication, corresponding to flow propagation)
  - Inflation (non-linear operation to increase heterogeneity)
- Complexity: $O(n^3)$ for an exact computation, $O(n)$ for an approximate one



Markov Cluster Algorithm (MCL) [van Dongen, 2000]

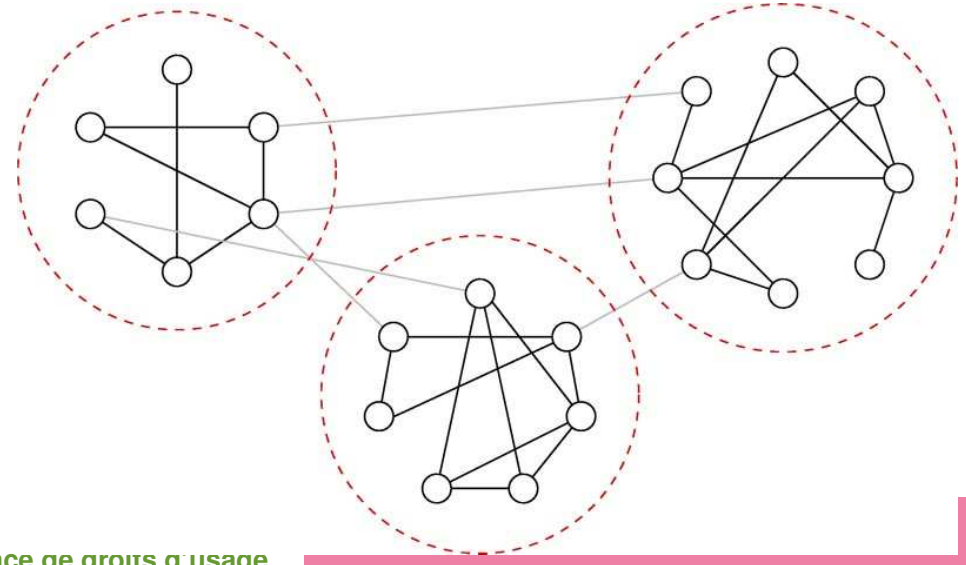
- Graph clustering algorithm
- Based as well on maximum flow simulation, in the whole graph
- Iteration of a matrix computation alternating:
  - Expansion (matrix multiplication, corresponding to flow propagation)
  - Inflation (non-linear operation to increase heterogeneity)
- Complexity: $O(n^3)$ for an exact computation, $O(n)$ for an approximate one



[van Dongen, 2000]

Deletion of the edges with the highest betwenness [Newman and Girvan, 2004]

- Top-down graph clustering algorithm
- Betwenness of an edge: number of minimal paths between two arbitrary vertices going through this edge
- General principle:
  1. Compute the betweenness of each edge in the graph
  2. Remove the edge with the highest betweenness
  3. Redo the whole process, betweenness computation included
- Complexity: $O(n^3)$ for a sparse graph



[Newman and Girvan, 2004]

Outline

Basics of Graph Theory

The Web as a graph

Social Networking Sites
- Typology
- Social Network Graphs
- Algorithms on social networks

Conclusion

Graph Mining on Other Graphs

- Graph mining tools used for the Web: also applicable to a variety of different graphs:
  - dictionary
  - encyclopedia
  - citation graph
  - ...
- Among these: social networking Web sites
- Attention: on undirected graphs, PageRank is proportional to degree! Not very useful.

Outline

Basics of Graph Theory

The Web as a graph

Social Networking Sites
- Typology
- Social Network Graphs
- Algorithms on social networks

Conclusion

Most popular social networks

Social networking sites the most popular in the world and in France (Web site rank with the most traffic, according to Alexa)

	World	France
SkyRock	51	3
YouTube	3	4
MySpace	17	7
Facebook	5	8
Dailymotion	61	11
EBay	18	12
Wikipedia	8	13
Meetic	565	27
ImageShack	47	53
hi5	15	59
Megavideo	133	80
Adult Friendfinder	55	82
Wat.tv	1568	88
Flickr	33	94
Orkut	19	>100
V Kontakte	28	>100
Friendster	39	>100

Typology of social networking sites

- Content-oriented
  - Cataloging (Books, Music, Links, Movies, Publications, Games)
  - Sharing (Images, Videos)
  - Edition
  - Sales
  - Discussion
- User-oriented
  - Pure social networks (Personal, Professional, Mixed)
  - Blogs
  - Dating

Outline

Basics of Graph Theory

The Web as a graph

Social Networking Sites
- Typology
- Social Network Graphs
- Algorithms on social networks

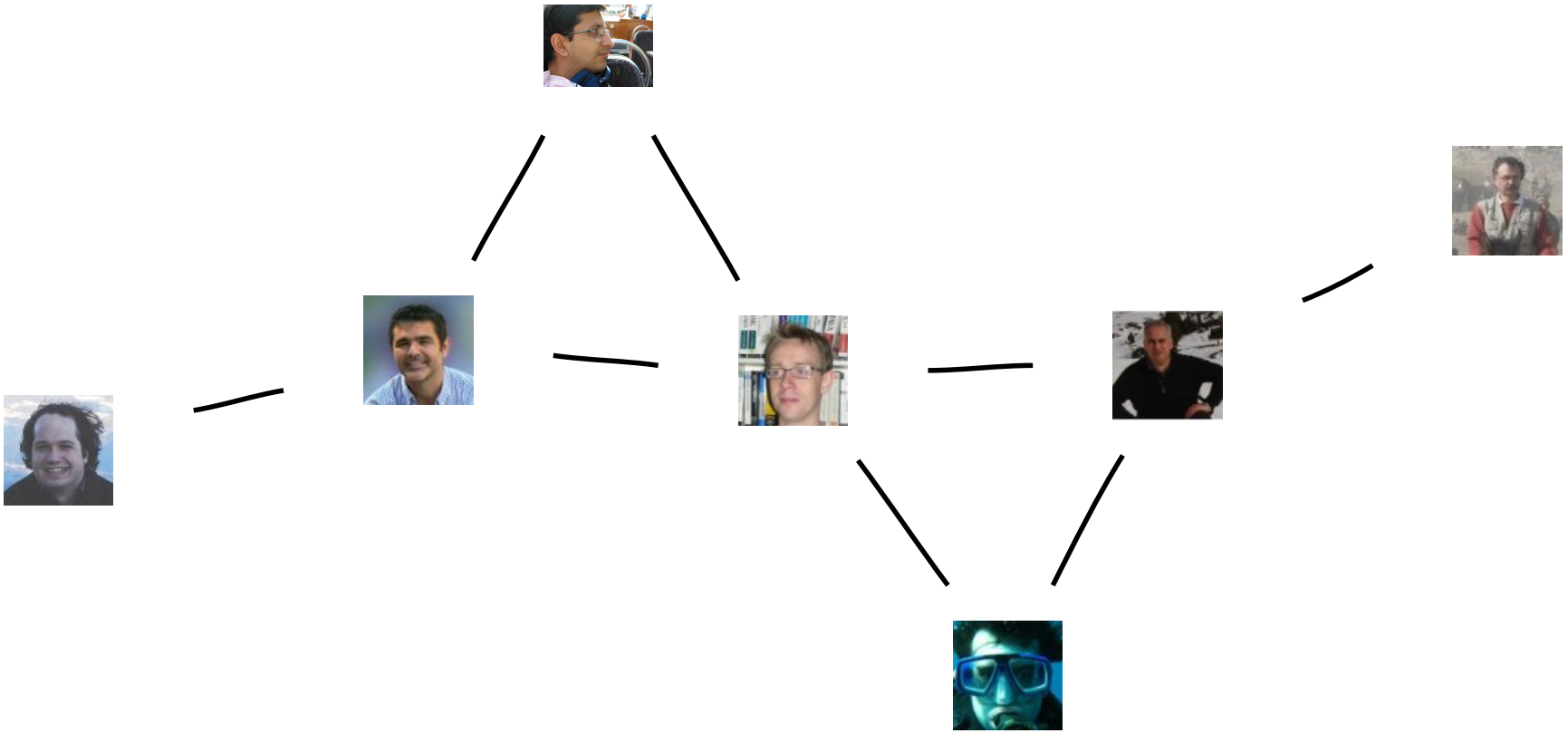
Conclusion

Graphs of Social Networks

- Natural model: social network = graph
- Entities = nodes, Relations = edges
- Depending on the specific case:
  - mostly undirected graphs
  - bipartite, *n*-partite
  - annotated edges, weighted edges

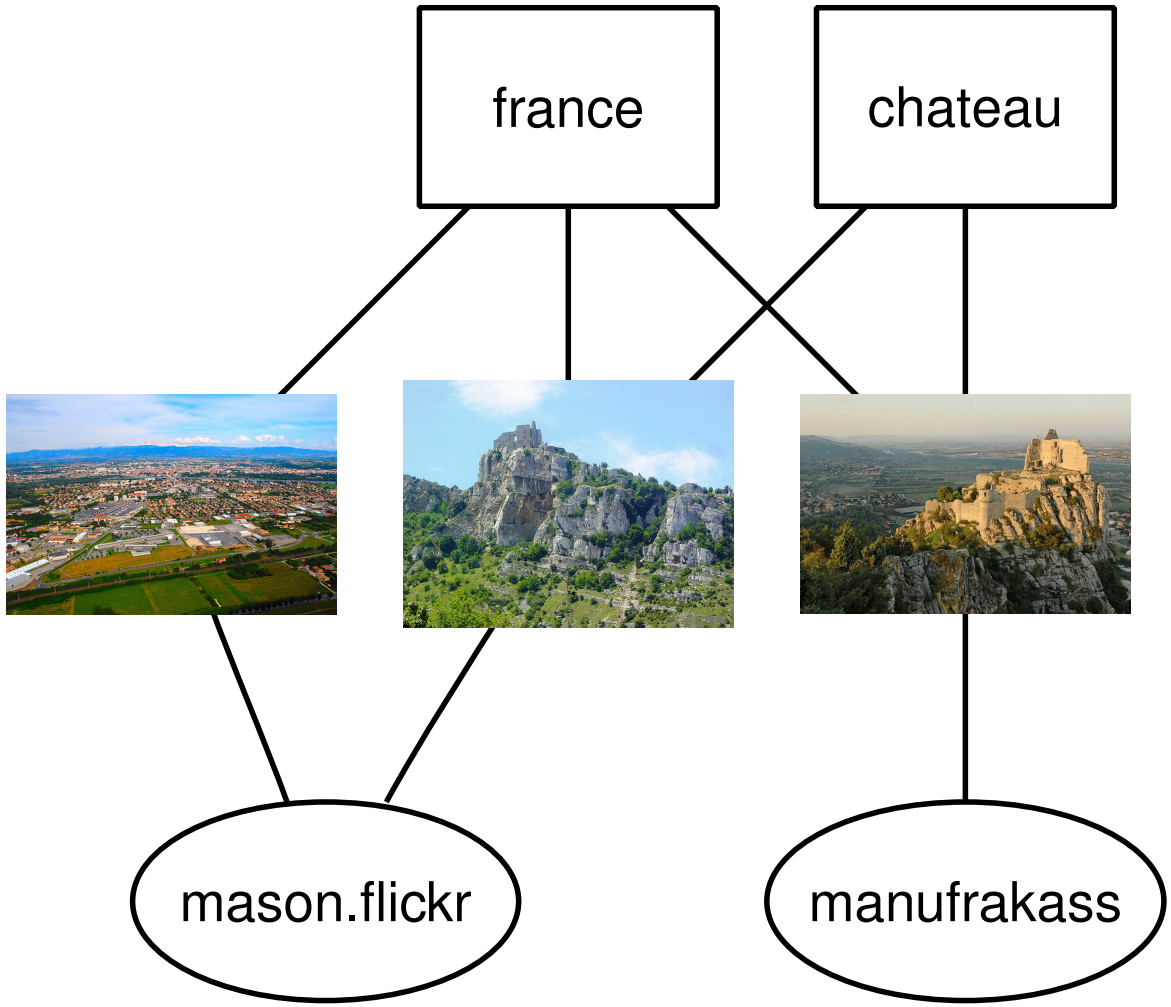
Undirected graph

Adapted for pure social networking sites with symmetrical relations (e.g., LinkedIn)



Multipartite graph

Adapted to most social networks for sharing content, with annotations, users, content, etc. (e.g., Flickr)



Six degrees of separation

- Idea that any two persons on Earth are separated by a chain of six persons such that any two successive links know each other
- Demonstrated by an experiment by Stanley Milgram [Travers and Milgram, 1969] (package to transmit to an unknown person, using acquaintances)
- Popularized in numerous media
- Number 6 is not too be taken too seriously! But main idea validated in more recent experiments
- In other fields:
  - Erdős number for scientific publications
  - Kevin Bacon for Hollywood movies

Common Characteristics of (most) social networks !

Characteristics of Social Networking Sites

Four important characteristics [Newman et al., 2006]:

Sparse graphs: much more edges than a complete graph

Small typical distance: shortest path between two nodes logarithmic with respect to the size of the graph

High clustering: if *a* is connected to *b* and *b* to *c*, then *b* has good probability of being linked to *c*

Power-law degree distribution: the number of nodes with degree *k* of the order of $k^{-\gamma}$ (γ constant)

More on this in Athens course on Collective Intelligence (TPT09)

Outline

Basics of Graph Theory

The Web as a graph

Social Networking Sites
- Typology
- Social Network Graphs
- Algorithms on social networks

Conclusion

Information retrieval with social score [Schenkel et al., 2008]

- Context: Multipartite graph, e.g., Flickr
- Purpose: bias results with one's social network
- Social weighting:
  - Given a friendship relation $F(u, u')$ (explicit or implicit) between two users, an extended friendship relation can be computed

$$\tilde{F}(u, u') = \frac{\alpha}{|U|} + (1 - \alpha) \max_{\text{chemin } u = u_0 \ldots u_k = u'} \prod_{i=0}^{k-1} F(u_i, u_{i+1})$$

  ($0 < \alpha < 1$ constant; $|U|$: number of users)
  - Instead of a global weighting

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, d)$$

  we choose a social weighting dependent of u:

$$\text{tf-idf}_u(t, d) = \left(\sum_{u' \in U} F(u, u') \cdot \text{tf}_{u'}(t, d) \right) \times \text{idf}(t, d)$$

Top-k with social score [Benedikt et al., 2008]

- Possible to adapt Fagin's trheshold algorithm...
- ... but impossible to precompute the scores tf-idf$_u(t, d)$ for each user
- To avoid too much computation time:
  1. Cluster the graph of users into strongly similar components
  2. Use the scores inside these components as estimates of the threshold in Fagin's algorithm
  3. $\Rightarrow$ gives approximate results, but good quality

Outline

Basics of Graph Theory

The Web as a graph

Social Networking Sites

Conclusion

Conclusion

What you should remember

- The Web seen as a graph
- PageRank, and its iterative computation
- Graph mining techniques applicable to a wide range of graphs
- Social networks share important common characteristics with the Web graph

To go further

- A good textbook [Chakrabarti, 2003]
- Two accessible influential research papers:
  - On HITS [Kleinberg, 1999]
  - On PageRank [Brin and Page, 1998]

Bibliography I

Serge Abiteboul, Mihai Preda, and Gregory Cobena. Adaptive on-line page importance computation. In *Proc. WWW*, May 2003.

Michael Benedikt, Sihem Amer Yahia, Laks Lakshmanan, and Julia Stoyanovich. Efficient network-aware search in collaborative tagging sites. In *Proc. VLDB*, Auckland, New Zealand, August 2008.

Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks*, 30(1–7): 107–117, April 1998.

Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer Networks*, 33(1-6): 309–320, 2000.

Soumen Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan Kaufmann, San Fransisco, USA, 2003.

Bibliography II

Andrew V. Goldberg and Robert E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM*, 35(4):921–940, October 1988.

Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 2004.

Mark Newman, Albert-László Barabási, and Duncan J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

Yann Ollivier and Pierre Senellart. Finding related pages using Green measures: An illustration with Wikipedia. In *Proc. AAAI*, Vancouver, Canada, July 2007.

Bibliography III

Ralf Schenkel, Tom Crecelius, Mouna Kacimi, Sebastian Michel, Thomas Neumann, Josiane X. Parreira, and Gerhard Weikum. Efficient top-k querying over social-tagging networks. In *Proc. SIGIR*, Singapore, Singapore, July 2008.

Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 34(4), December 1969.

Stijn Marinus van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.

Licence de droits d'usage

Contexte public } avec modifications

Par le téléchargement ou la consultation de ce document, l'utilisateur accepte la licence d'utilisation qui y est attachée, telle que détaillée dans les dispositions suivantes, et s'engage à la respecter intégralement.

La licence confère à l'utilisateur un droit d'usage sur le document consulté ou téléchargé, totalement ou en partie, dans les conditions définies ci-après et à l'exclusion expresse de toute utilisation commerciale.
Le droit d'usage défini par la licence autorise un usage à destination de tout public qui comprend :
– le droit de reproduire tout ou partie du document sur support informatique ou papier,
– le droit de diffuser tout ou partie du document au public sur support papier ou informatique, y compris par la mise à la disposition du public sur un réseau numérique,
– le droit de modifier la forme ou la présentation du document,
– le droit d'intégrer tout ou partie du document dans un document composite et de le diffuser dans ce nouveau document, à condition que :
  – L'auteur soit informé.

Les mentions relatives à la source du document et/ou à son auteur doivent être conservées dans leur intégralité.

Le droit d'usage défini par la licence est personnel et non exclusif.
Tout autre usage que ceux prévus par la licence est soumis à autorisation préalable et expresse de l'auteur : sitepedago@telecom-paristech.fr