Cascades in networks: a simple theory and applications

Graduate Summer School:
Games and Contracts for Cyber-Physical Security
IPAM, UCLA

21 July 2015 (4:30-5:15)

Yongwhan Lim, Asuman Ozdaglar, Alex Teytelboym

MIT, MIT, Oxford

A simple model of cascades

Motivation

- Many cascade phenomena that occur in social, economic, and physical networks are **irreversible** (at least temporarily):
  - positive: innovation/technology adoption, social platform use, mobile phone contracts etc.
  - negative: spread of incurable diseases, bank failures, outages in power grids, drug addiction, dropping out of high school etc.
- We call any such irreversible change a **switch**.

Motivation

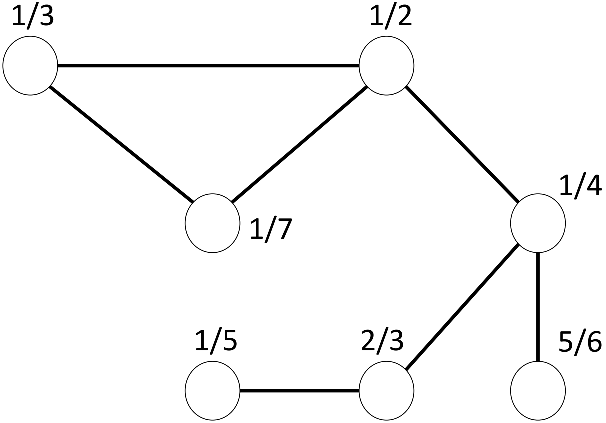
- Many cascade phenomena that occur in social, economic, and physical networks are **irreversible** (at least temporarily):
  - positive: innovation/technology adoption, social platform use, mobile phone contracts etc.
  - negative: spread of incurable diseases, bank failures, outages in power grids, drug addiction, dropping out of high school etc.
- We call any such irreversible change a **switch**.

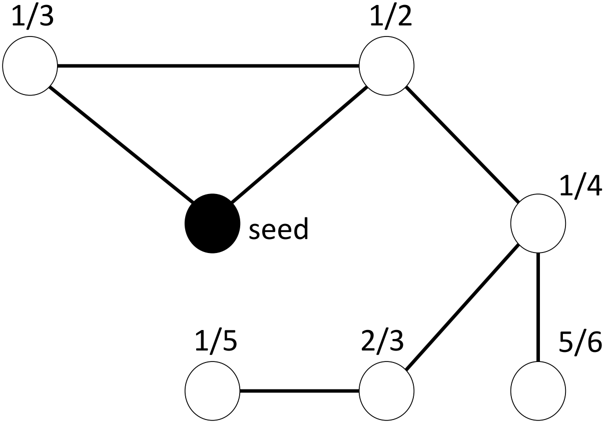
Motivation

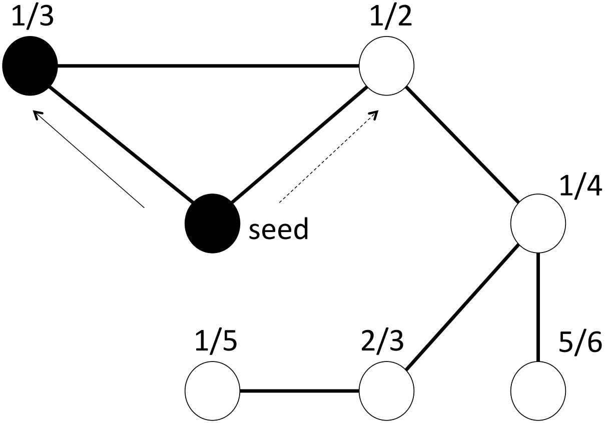
- These cascade phenomena:
  - exhibit **path dependence** - initial conditions matter.
  - exhibit **network effects** - agents are heterogeneously affected by their neighbors - so **network structure** matters.

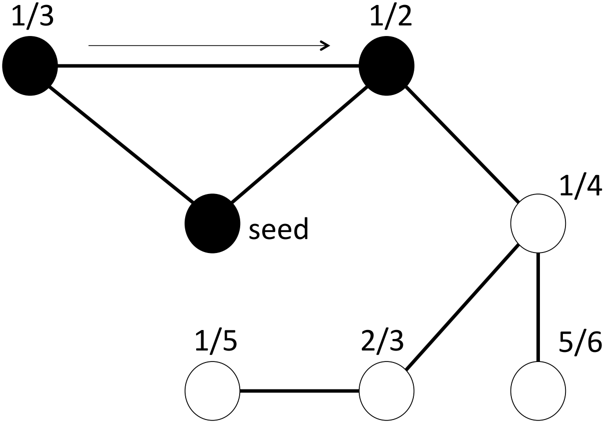
Typical model

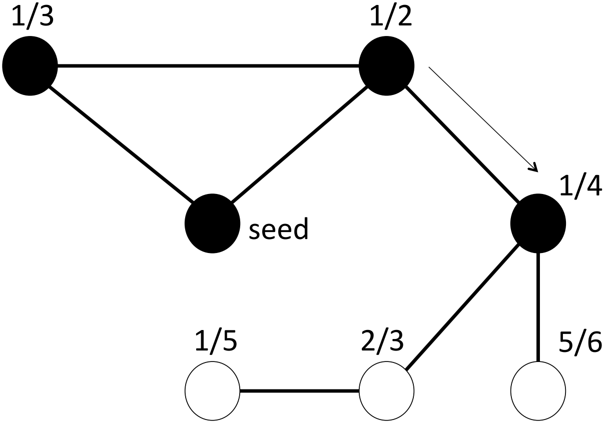
- Granovetter's familiar **linear threshold model** on networks captures all these features:
  - Initially, all agents in a network are in their default state.
  - Then, some agents ("seeds") are switched.
  - Subsequent agents switch if the **proportion** of their neighbors who have switched exceeds some individual threshold.
  - Once an agent switches, he is switched forever.

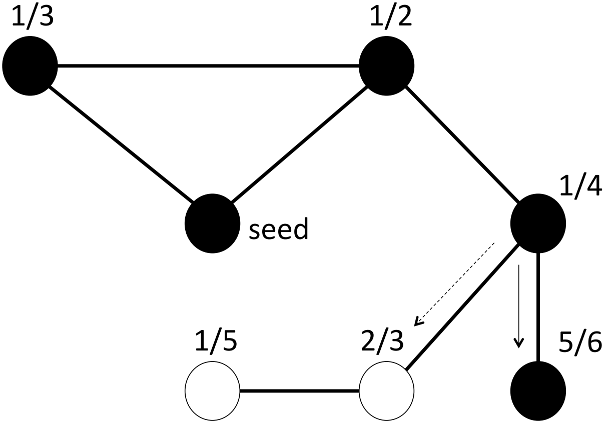

1/3
1/2
1/4
1/7
1/5
2/3
5/6

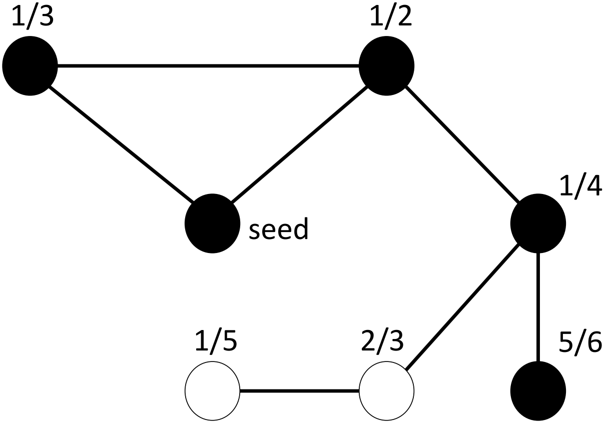

1/3
1/2
1/4
seed
1/5
2/3
5/6


1/3
1/2
1/4
seed
1/5
2/3
5/6


1/3
1/2
1/4
seed
1/5
2/3
5/6


1/3
1/2
1/4
seed
1/5
2/3
5/6


1/3
1/2
seed
1/4
1/5
2/3
5/6


1/3
1/2
seed
1/4
1/5
2/3
5/6

Motivation

- This model is useful, but notoriously difficult to analyse (size $2^{\binom{n}{2}} \times 2^n$ with n agents). Three multidimensional parameters:
  - network topology
  - each agent's threshold
  - initial seeds

Previous work vs. this talk

- Necessary and sufficient condition for *complete contagion* for regular infinite lattices with single seeds using *cohesive sets* (Morris, 2000). We cover general graphs, any cascade size and arbitrary seed sets.
- Complete characterization of the *switch set* also in terms of cohesive sets (Acemoglu et al., 2011). More clustering → fewer switches. We consider the expected number of switches and show that general comparative statics do not depend on macroscopic properties of graphs.

Previous work vs. this talk

- Necessary and sufficient condition for *complete contagion* for regular infinite lattices with single seeds using *cohesive sets* (Morris, 2000). We cover general graphs, any cascade size and arbitrary seed sets.
- Complete characterization of the *switch set* also in terms of cohesive sets (Acemoglu et al., 2011). More clustering $\rightarrow$ fewer switches. We consider the expected number of switches and show that general comparative statics do not depend on macroscopic properties of graphs.

Previous work vs. this talk

- Necessary and sufficient condition for *complete contagion* for regular infinite lattices with single seeds using *cohesive sets* (Morris, 2000). We cover general graphs, any cascade size and arbitrary seed sets.
- Complete characterization of the *switch set* also in terms of cohesive sets (Acemoglu et al., 2011). More clustering $\rightarrow$ fewer switches. We consider the expected number of switches and show that general comparative statics do not depend on macroscopic properties of graphs.

Previous work vs. this talk

- Necessary and sufficient condition for *complete contagion* for regular infinite lattices with single seeds using *cohesive sets* (Morris, 2000). We cover general graphs, any cascade size and arbitrary seed sets.
- Complete characterization of the *switch set* also in terms of cohesive sets (Acemoglu et al., 2011). More clustering $\rightarrow$ fewer switches. We consider the expected number of switches and show that general comparative statics do not depend on macroscopic properties of graphs.

Previous work vs. this talk

- Algorithms for choosing a seed set to maximize or minimize switches (Kempe et al., 2003; Blume et al., 2011). We focus on network design rather than on seed set selection.
- Evolutionary models (ergodic Markov chains) where agents can switch back and forth (Young, 2006). Our process is progressive/monotonic i.e. initial conditions matter.
- For some applications, such as complete contagion, path-dependent and ergodic models are equivalent (Adam et al., 2013).

Previous work vs. this talk

- Algorithms for choosing a seed set to maximize or minimize switches (Kempe et al., 2003; Blume et al., 2011). We focus on network design rather than on seed set selection.
- Evolutionary models (ergodic Markov chains) where agents can switch back and forth (Young, 2006). Our process is progressive/monotonic i.e. initial conditions matter.
- For some applications, such as complete contagion, path-dependent and ergodic models are equivalent (Adam et al., 2013).

Previous work vs. this talk

- Algorithms for choosing a seed set to maximize or minimize switches (Kempe et al., 2003; Blume et al., 2011). We focus on network design rather than on seed set selection.
- Evolutionary models (ergodic Markov chains) where agents can switch back and forth (Young, 2006). Our process is progressive/monotonic i.e. initial conditions matter.
- For some applications, such as complete contagion, path-dependent and ergodic models are equivalent (Adam et al., 2013).

Previous work vs. this talk

- Algorithms for choosing a seed set to maximize or minimize switches (Kempe et al., 2003; Blume et al., 2011). We focus on network design rather than on seed set selection.
- Evolutionary models (ergodic Markov chains) where agents can switch back and forth (Young, 2006). Our process is progressive/monotonic i.e. initial conditions matter.
- For some applications, such as complete contagion, path-dependent and ergodic models are equivalent (Adam et al., 2013).

Previous work vs. this talk

- Algorithms for choosing a seed set to maximize or minimize switches (Kempe et al., 2003; Blume et al., 2011). We focus on network design rather than on seed set selection.
- Evolutionary models (ergodic Markov chains) where agents can switch back and forth (Young, 2006). Our process is progressive/monotonic i.e. initial conditions matter.
- For some applications, such as complete contagion, path-dependent and ergodic models are equivalent (Adam et al., 2013).

Outline of this talk

- We develop a tractable model of cascades in networks.
- We introduce a new centrality concept called **cascade centrality**.
- We characterize the expected number of switches using cascade centrality in various classes graphs.

Model

Model: preliminaries

- Simple, undirected graph $G(V, E)$ with a set of n agents $V := \{1, \ldots, n\}$ and a set of m links E.
- Neighbors of $i \in V$ denoted $N_i(G) := \{j | (j, i) \in E\}$ and the degree of i as $d_i := |N_i(G)|$.
- A *threshold* for agent i is a random variable Θ_i drawn from a probability distribution with support $[0, 1]$.
- The associated multivariate probability distribution for all the nodes in the graph is $f(\theta)$.
- Each agent is $i \in V$ assigned a threshold θ_i. Let's define the threshold profile of agents as $\boldsymbol{\theta} := (\theta_i)_{i \in V}$. A **network** G_θ is a graph endowed with a threshold profile.

Model: preliminaries

- Simple, undirected graph $G(V, E)$ with a set of n agents $V := \{1, \ldots, n\}$ and a set of m links E.
- Neighbors of $i \in V$ denoted $N_i(G) := \{j | (j, i) \in E\}$ and the degree of i as $d_i := |N_i(G)|$.
- A *threshold* for agent i is a random variable Θ_i drawn from a probability distribution with support $[0, 1]$.
- The associated multivariate probability distribution for all the nodes in the graph is $f(\theta)$.
- Each agent is $i \in V$ assigned a threshold θ_i. Let's define the threshold profile of agents as $\boldsymbol{\theta} := (\theta_i)_{i \in V}$. A **network** G_θ is a graph endowed with a threshold profile.

Model: preliminaries

- Simple, undirected graph $G(V, E)$ with a set of n agents $V := \{1, \ldots, n\}$ and a set of m links E.
- Neighbors of $i \in V$ denoted $N_i(G) := \{j | (j, i) \in E\}$ and the degree of i as $d_i := |N_i(G)|$.
- A *threshold* for agent i is a random variable Θ_i drawn from a probability distribution with support $[0, 1]$.
- The associated multivariate probability distribution for all the nodes in the graph is $f(\theta)$.
- Each agent is $i \in V$ assigned a threshold θ_i. Let's define the threshold profile of agents as $\boldsymbol{\theta} := (\theta_i)_{i \in V}$. A **network** G_θ is a graph endowed with a threshold profile.

Model: preliminaries

- Simple, undirected graph $G(V, E)$ with a set of n agents $V := \{1, \dots, n\}$ and a set of m links E.
- Neighbors of $i \in V$ denoted $N_i(G) := \{j | (j, i) \in E\}$ and the degree of i as $d_i := |N_i(G)|$.
- A *threshold* for agent i is a random variable Θ_i drawn from a probability distribution with support $[0, 1]$.
- The associated multivariate probability distribution for all the nodes in the graph is $f(\boldsymbol{\theta})$.
- Each agent is $i \in V$ assigned a threshold θ_i. Let's define the threshold profile of agents as $\boldsymbol{\theta} := (\theta_i)_{i \in V}$. A **network** G_θ is a graph endowed with a threshold profile.

Model: preliminaries

- Simple, undirected graph $G(V, E)$ with a set of n agents $V := \{1, \ldots, n\}$ and a set of m links E.
- Neighbors of $i \in V$ denoted $N_i(G) := \{j | (j, i) \in E\}$ and the degree of i as $d_i := |N_i(G)|$.
- A *threshold* for agent i is a random variable Θ_i drawn from a probability distribution with support $[0, 1]$.
- The associated multivariate probability distribution for all the nodes in the graph is $f(\boldsymbol{\theta})$.
- Each agent is $i \in V$ assigned a threshold θ_i. Let's define the threshold profile of agents as $\boldsymbol{\theta} := (\theta_i)_{i \in V}$. A **network** $G_{\boldsymbol{\theta}}$ is a graph endowed with a threshold profile.

Model: dynamics

- At time $t = 0$, a subset of agents $S_0 \subseteq V$ is selected is a seed set. We assume that at $t = 0$ agents switch if and only if they are in the seed set.
- For any $t \geq 0$ and any $i \in V \setminus S_0(G_\theta)$:

$$\frac{|S_0(G_\theta) \cap N_i(G_\theta)|}{|N_i(G_\theta)|} \geq \theta_i \Rightarrow i \in S_1(G_\theta)$$

Model: dynamics

- At time $t = 0$, a subset of agents $S_0 \subseteq V$ is selected is a seed set. We assume that at $t = 0$ agents switch if and only if they are in the seed set.
- For any $t \geq 0$ and any $i \in V \setminus S_0(G_\theta)$:

$$\frac{|S_0(G_\theta) \cap N_i(G_\theta)|}{|N_i(G_\theta)|} \geq \theta_i \Rightarrow i \in S_1(G_\theta)$$

Model: dynamics

- At time $t = 0$, a subset of agents $S_0 \subseteq V$ is selected is a seed set. We assume that at $t = 0$ agents switch if and only if they are in the seed set.
- For any $t \geq 0$ and any $i \in V \setminus S_0(G_{\boldsymbol{\theta}})$:

$$\frac{|S_0(G_{\boldsymbol{\theta}}) \cap N_i(G_{\boldsymbol{\theta}})|}{|N_i(G_{\boldsymbol{\theta}})|} \geq \theta_i \Rightarrow i \in S_1(G_{\boldsymbol{\theta}})$$

Model

- This means that any agent who has not switched by some period t, switches in time period $t+1$ if the proportion of his neighbors who switched is greater or equal to his threshold θ_i. For a given period $t \geq 0$ and node $i \in V \setminus \cup_{\tau=0}^{t-1} S_\tau(G_{\boldsymbol{\theta}})$ will switch at t if

$$\frac{|\{\cup_{\tau=0}^{t-1} S_\tau(G_{\boldsymbol{\theta}})\} \cap N_i(G_{\boldsymbol{\theta}})|}{|N_i(G_{\boldsymbol{\theta}})|} \geq \theta_i \Rightarrow i \in S_t(G_{\boldsymbol{\theta}})$$

For a given network $G_{\boldsymbol{\theta}}$, define the fixed point of the process as $S(G_{\boldsymbol{\theta}}, S_0)$ s.t. $S = S_0(G_{\boldsymbol{\theta}}) \Rightarrow S_t(G_{\boldsymbol{\theta}}) = \emptyset$ for all $t > 0$.

Model

- Let's fix a seed S_0 and a graph G, and re-run the process by drawing the agents' thresholds from $f(\boldsymbol{\theta})$ each time.
- The expected probability of agent i switching is:

$$\mathbb{P}_i(G, S_0) = \int_{\mathbb{R}^n} |S(G_{\theta}, S_0) \cap \{i\}| f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Total expected number of switches in graph G with seed S_0 is:

$$\mathbb{E}[S(G, S_0)] = \int_{\mathbb{R}^n} |S(G_{\theta}, S_0)| f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sum_{i=1}^{n} \mathbb{P}_i(G, S_0)$$

Model

- Let's fix a seed S_0 and a graph G, and re-run the process by drawing the agents' thresholds from $f(\boldsymbol{\theta})$ each time.
- The expected probability of agent i switching is:

$$\mathbb{P}_i(G, S_0) = \int_{\mathbb{R}^n} |S(G_{\boldsymbol{\theta}}, S_0) \cap \{i\}| f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Total expected number of switches in graph G with seed S_0 is:

$$\mathbb{E}[S(G, S_0)] = \int_{\mathbb{R}^n} |S(G_{\theta}, S_0)| f(\theta) d\boldsymbol{\theta} = \sum_{i=1}^{n} \mathbb{P}_i(G, S_0)$$

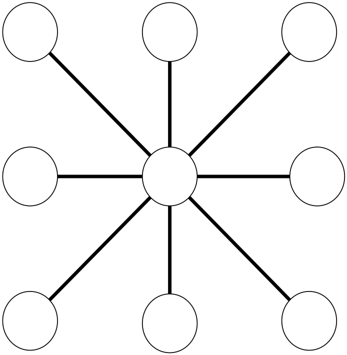
Model

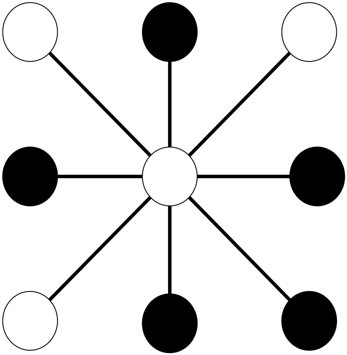
- Let's fix a seed S_0 and a graph G, and re-run the process by drawing the agents' thresholds from $f(\boldsymbol{\theta})$ each time.
- The expected probability of agent i switching is:

$$\mathbb{P}_i(G, S_0) = \int_{\mathbb{R}^n} |S(G_{\boldsymbol{\theta}}, S_0) \cap \{i\}| f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Total expected number of switches in graph G with seed S_0 is:

$$\mathbb{E}[S(G, S_0)] = \int_{\mathbb{R}^n} |S(G_{\boldsymbol{\theta}}, S_0)| f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sum_{i=1}^{n} \mathbb{P}_i(G, S_0)$$





Uniform thresholds

Lemma

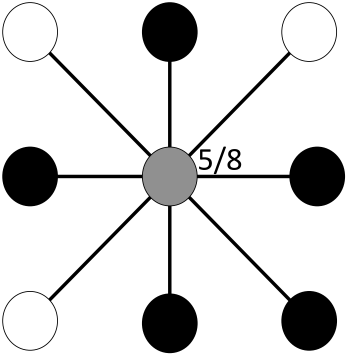
Let $\{G(n)\}_{n \in \mathbb{N}^+}$ be set of star networks of orders $n \in \mathbb{N}^+$ in which i is a center and the seed set is $S_0 \subseteq V \setminus \{i\}$, then

$$\mathbb{P}_i(G(n), S_0) = \frac{|S_0|}{d_i(G(n))}$$

for almost all G_n if and only if $\Theta_i \sim \mathcal{U}[0, 1]$.

Moreover, we can prove that

$$\mathbb{P}_i(G, S_0) = \sum_{j \in N_i(G)} \frac{\mathbb{P}_j(G | i \notin S)}{d_i}$$


5/8

Uniform thresholds

Assumption

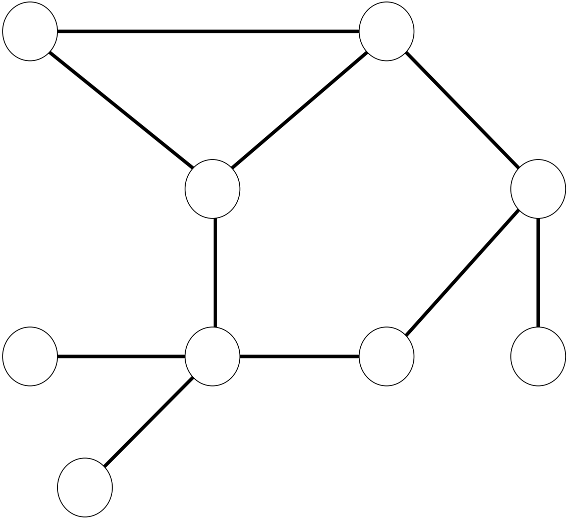
For any $G_{\boldsymbol{\theta}}$ and every $i \in V$, $\Theta_i \sim \mathcal{U}(0,1)$ and independent.

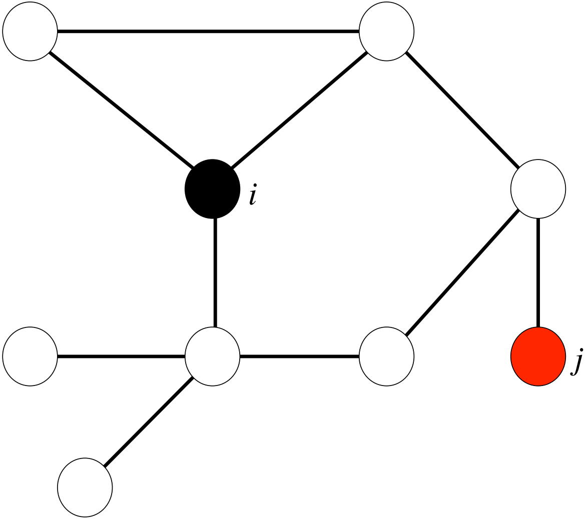
It's the Laplacian prior and not actually a very restrictive assumption.

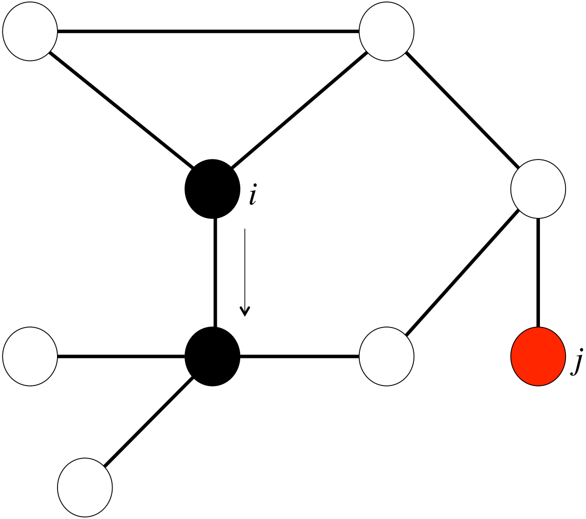
Paths

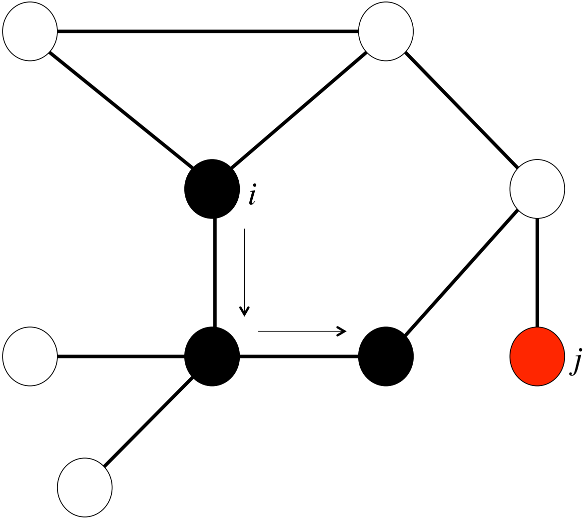
Definition

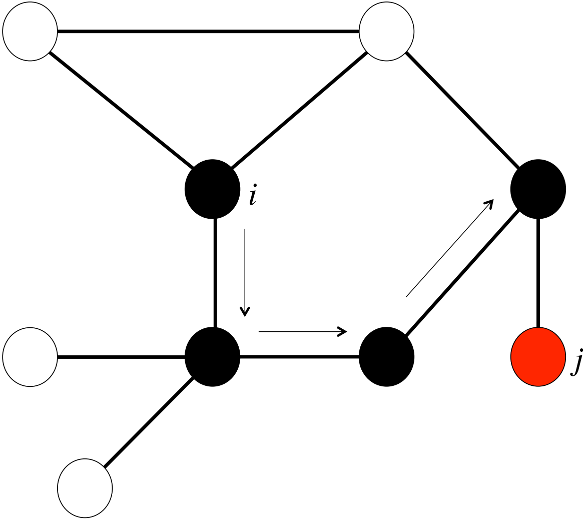
A sequence of nodes $P = (i_0, \cdots, i_k)$ on a graph G is a path if $i_j \in N_{i_{j-1}}(G)$ for all $1 \leq j \leq k$ and each $i_j \in P$ is distinct.

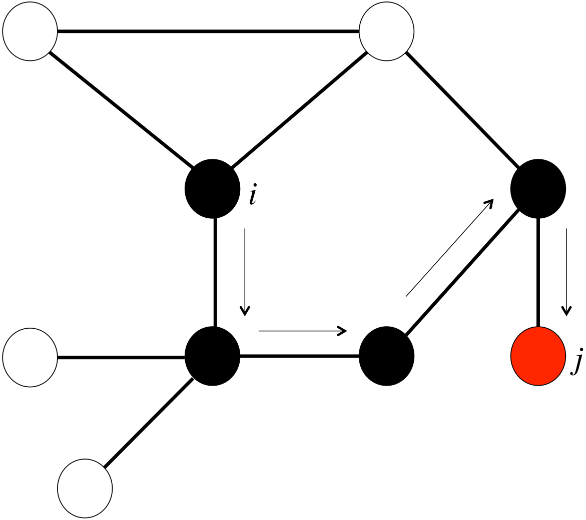


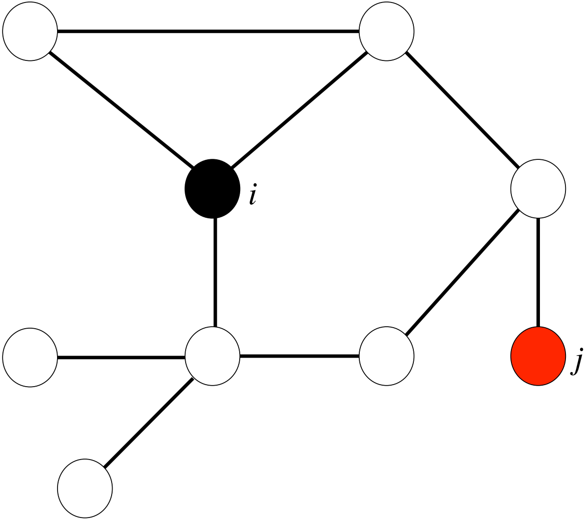

i
j

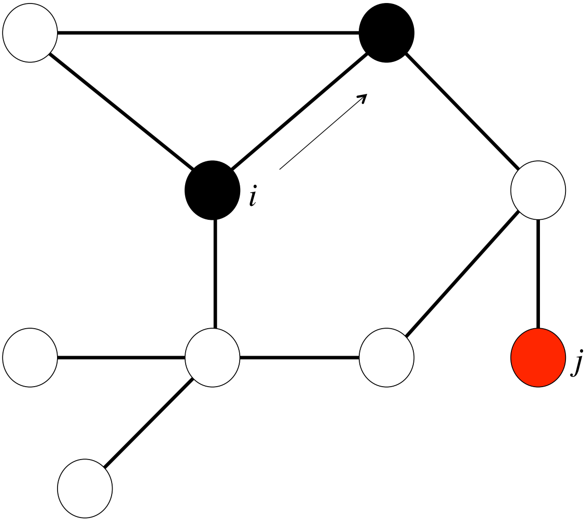

i
j

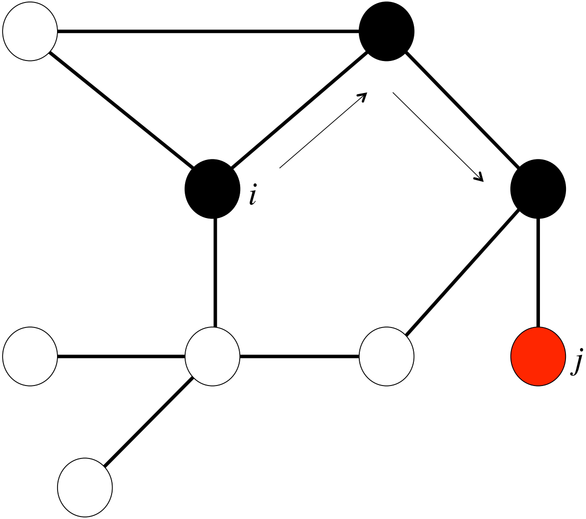

i
j

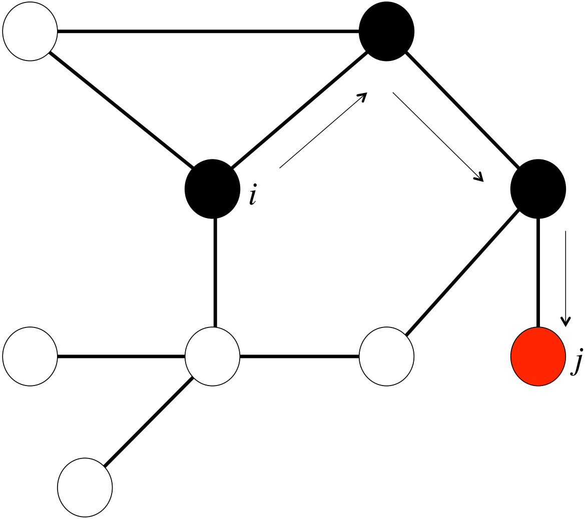

i
j

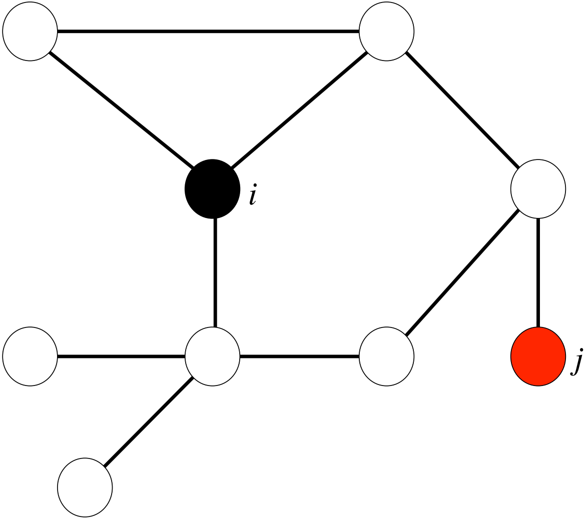

i
j


i
j

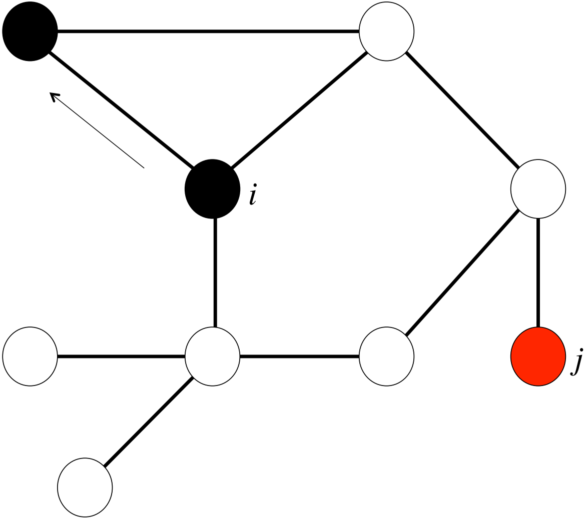

i
j

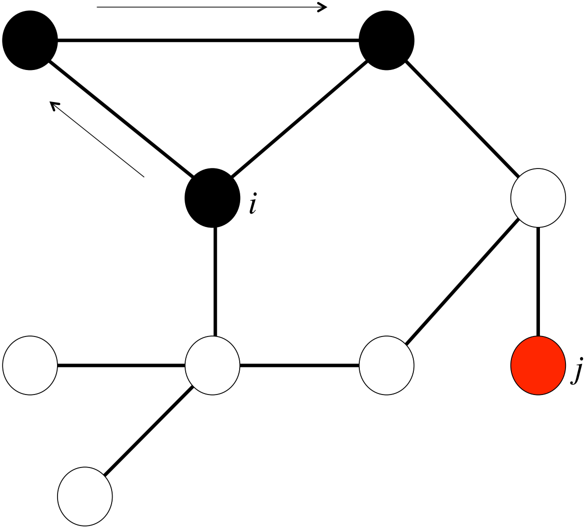

i
j


i
j

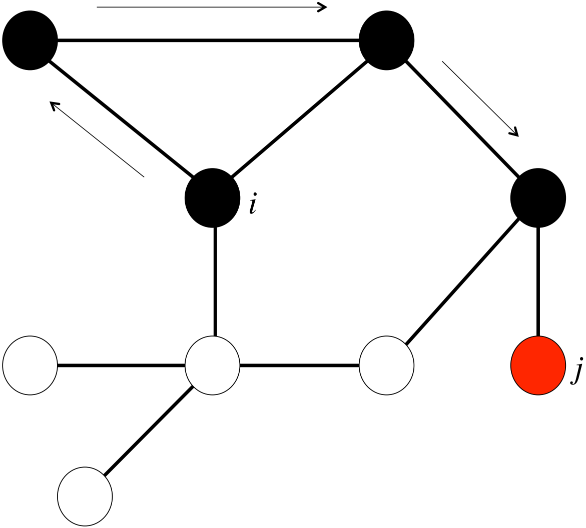


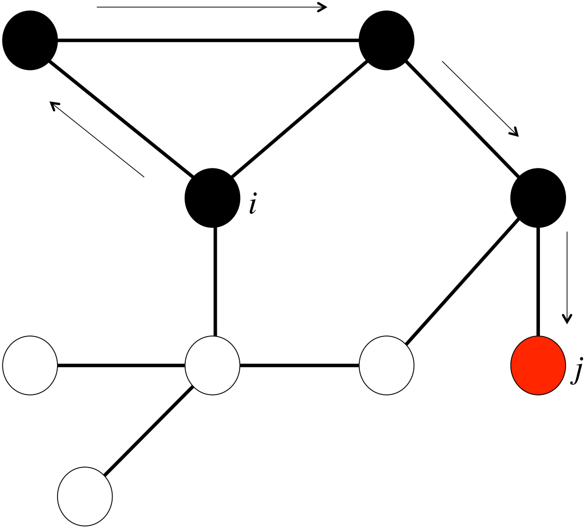
i
j


i
j



i
j


i
j


i
j

Degree sequence product

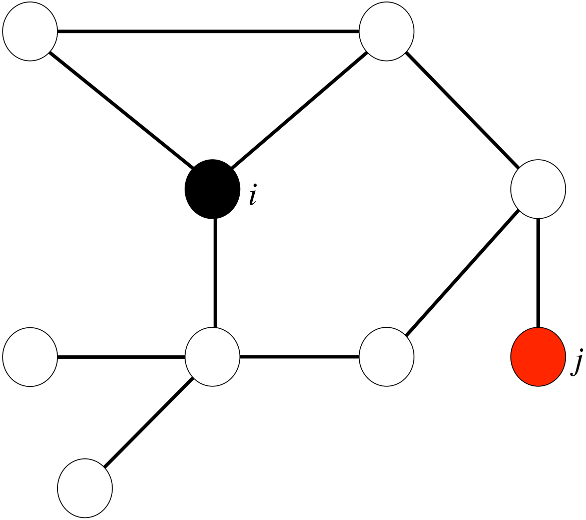
Definition

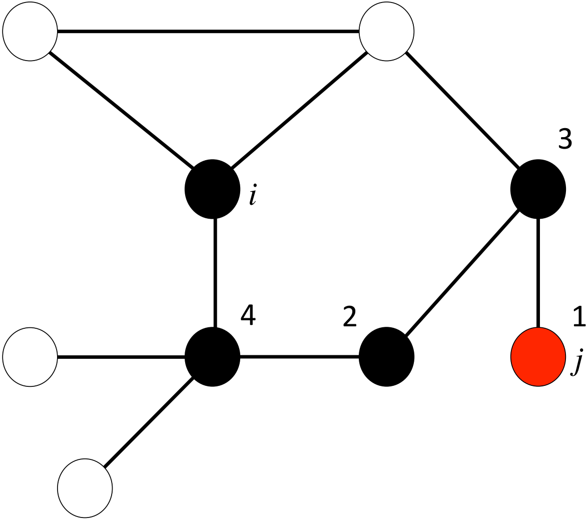
For a path P, a *degree sequence* along any path P is $(d_i(G))_{i \in P \setminus \{i_0\}}$.

Definition

A *degree sequence product* along P is:

$$\chi_P := \prod_{i \in P \setminus \{i_0\}} d_i(G)$$


i
j


3
i
4
2
1
j

Key proposition

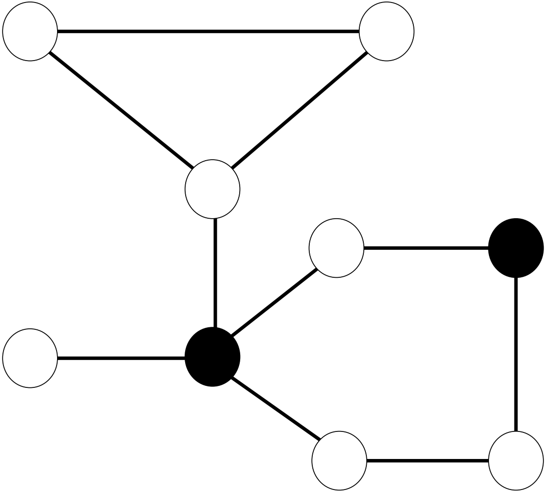
For any G and S_0, let $\mathcal{P}_{ji}$ be the set of all paths beginning at $j \in S_0$ and ending at $i \in V \setminus S_0$ and $\mathcal{P}^*_{ji} \subseteq \mathcal{P}_{ji}$ denote the subset of those paths that exclude any other node in S_0.

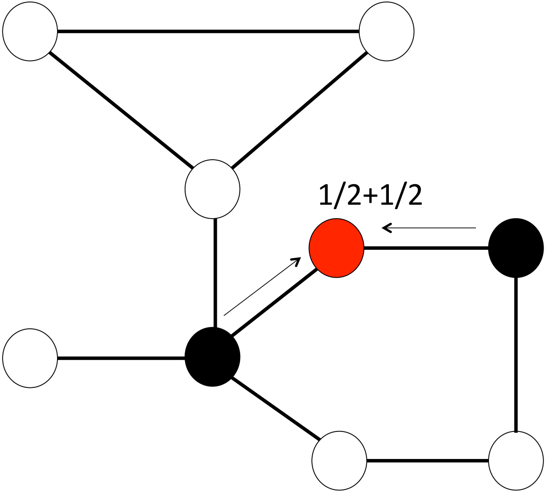
Proposition

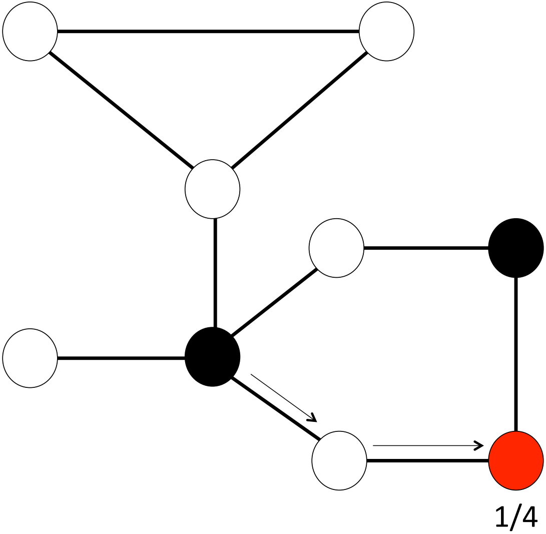
Given a graph G and seed S_0, the probability that node $i \in V \setminus S_0$ switches is:

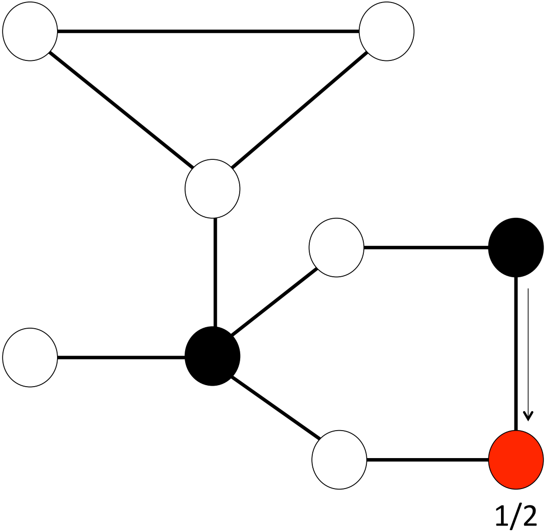
$$\mathbb{P}_i(G, S_0) = \sum_{j \in S_0} \sum_{P \in \mathcal{P}^*_{ji}} \frac{1}{\chi_P}$$

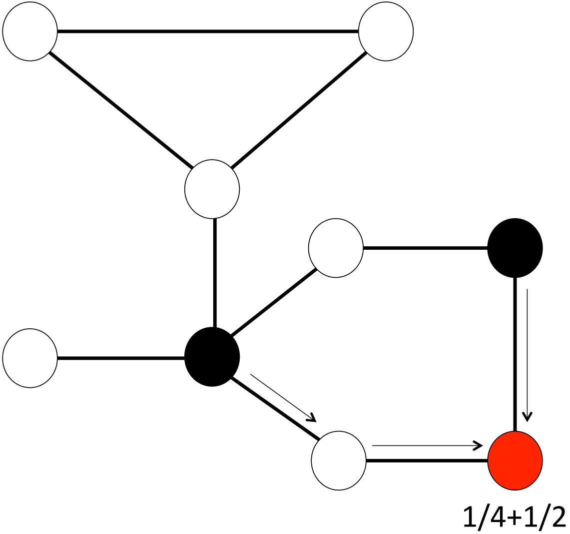
See Kempe et al. (2003).

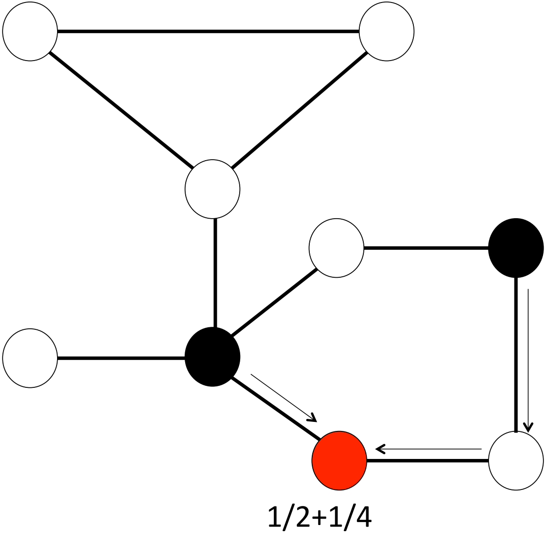


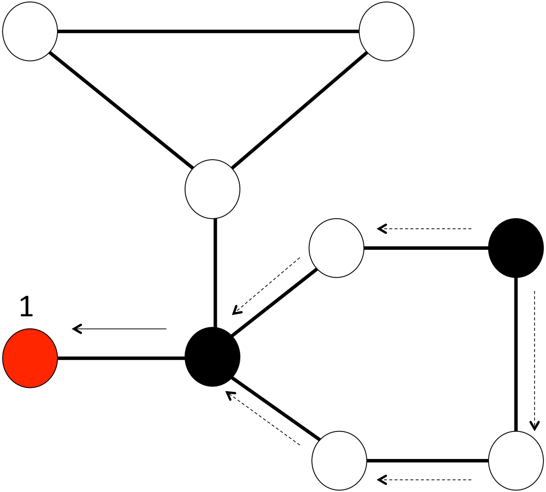

1/2+1/2

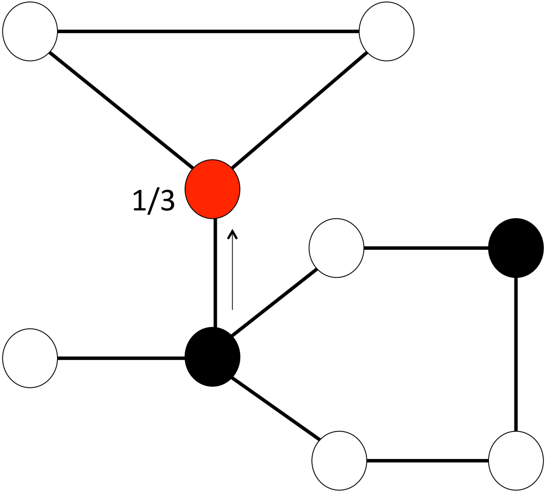

1/4

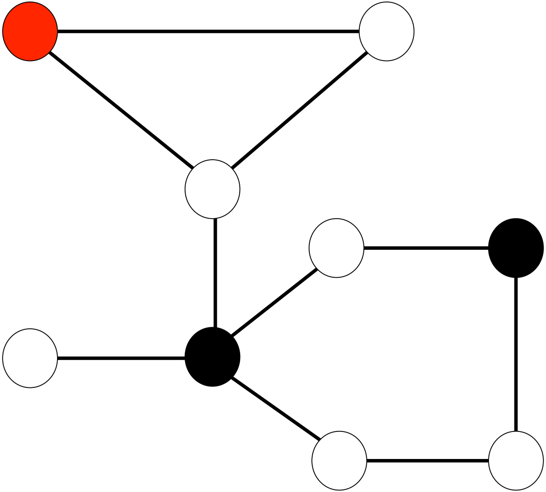

1/2

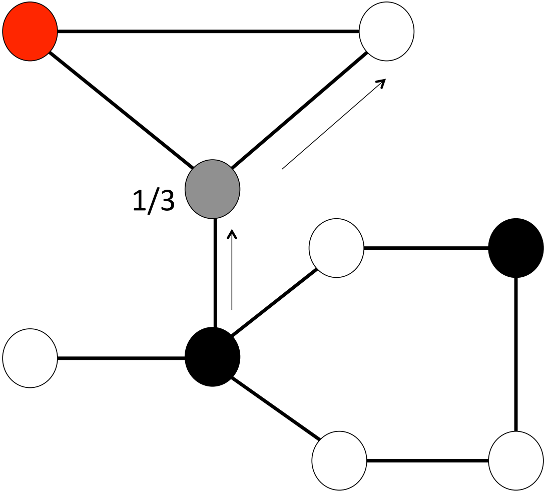

1/4+1/2

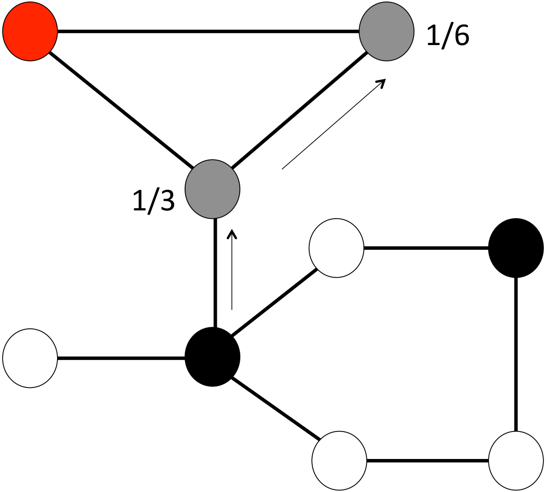

1/2+1/4

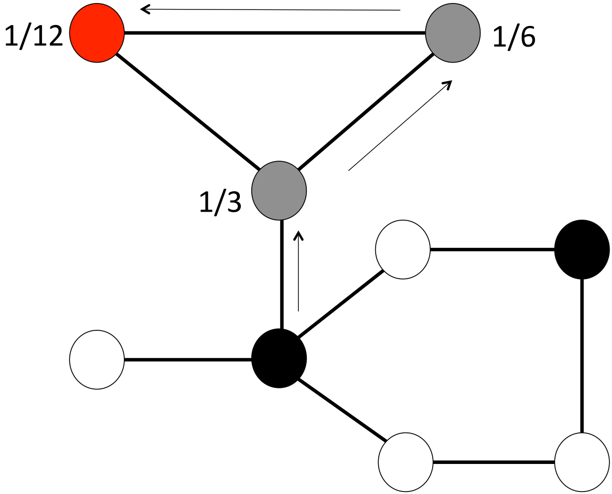

1

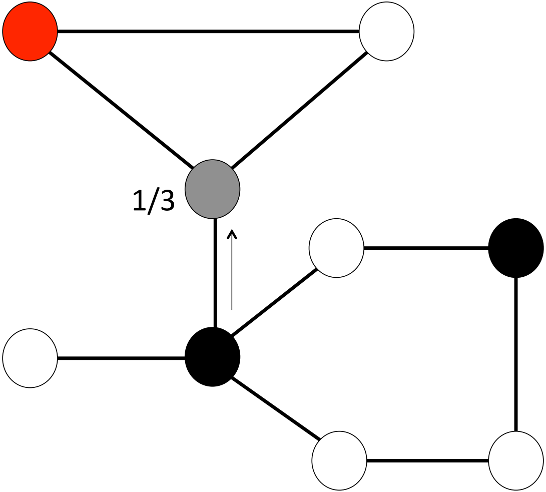

1/3

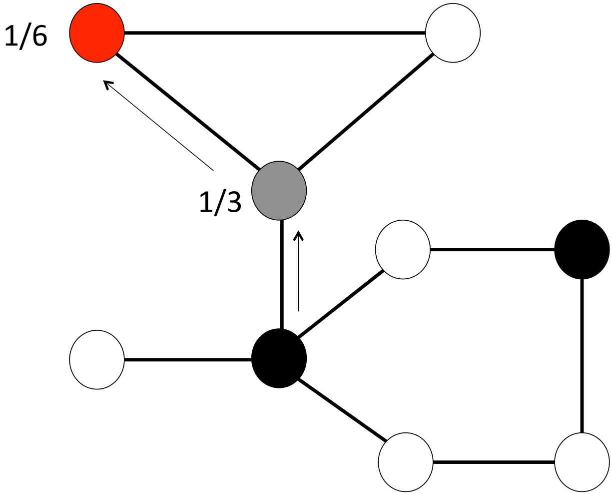


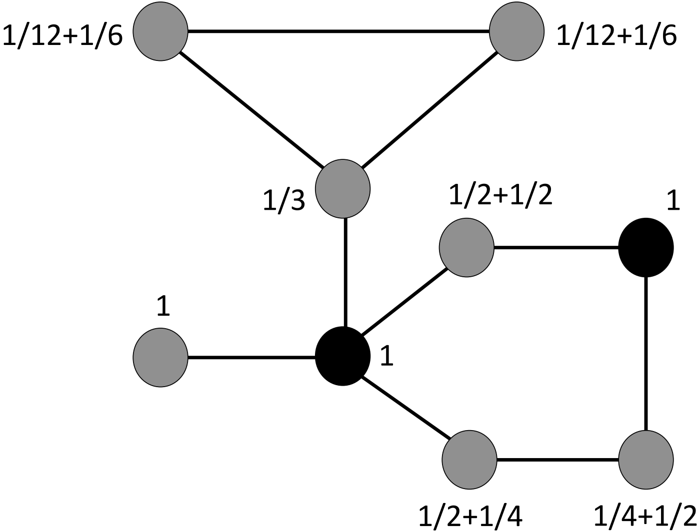

1/3


1/6
1/3


1/12
1/6
1/3


1/3


1/6
1/3


1/12+1/6
1/12+1/6
1/3
1/2+1/2
1
1
1
1/2+1/4
1/4+1/2

Cascade centrality

Definition

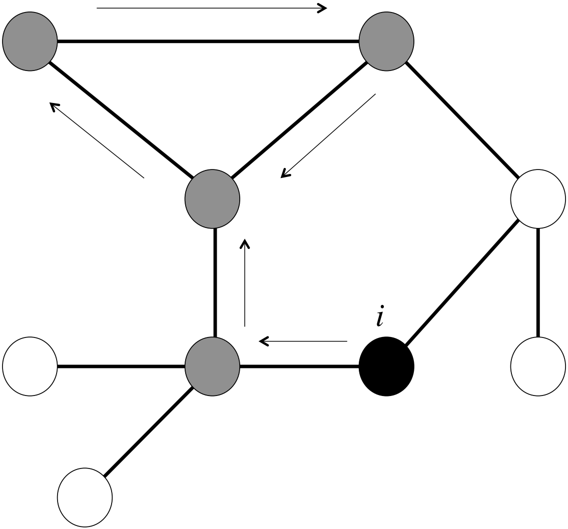
Cascade centrality of node i in graph G is the expected number of switches in that graph given i is the seed, namely

$$\mathcal{C}_i(G) := \mathbb{E}[S(G,\{i\})] = 1 + \sum_{j \in V \setminus \{i\}} \mathbb{P}_j(G,\{i\}) = 1 + \sum_{j \in V \setminus \{i\}} \sum_{P \in \mathcal{P}_{ij}} \frac{1}{\chi_P}$$

Loops

Definition

A sequence of nodes $L = (i_0, \ldots, i_k)$ on a graph G is a loop if $(i_0, \ldots, i_{k-1})$ is a path and $i_k \in \{i_0, \ldots, i_{k-2}\}$ for some $k \geq 2$.


i

Cascade centrality

Theorem

The cascade centrality of any node i in G is:

$$\mathcal{C}_i(G) = 1 + d_i - \sum_{j \in V \setminus \{i\}} \sum_{L \in \mathcal{L}_{ij}} \frac{1}{\chi_L}$$

where χ_L is the degree sequence product along a loop.

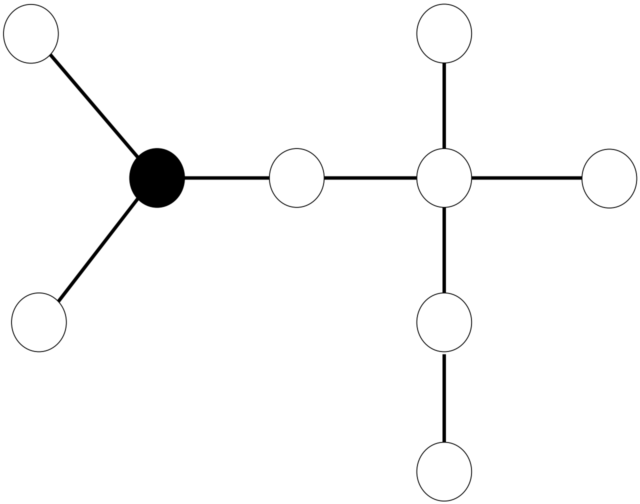
Analytical Results

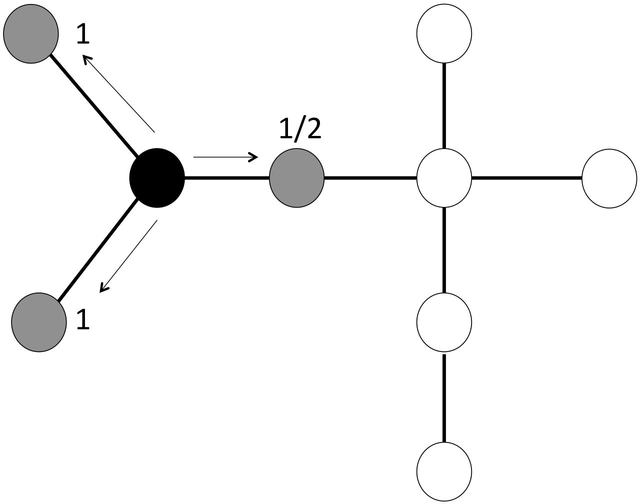
Tree

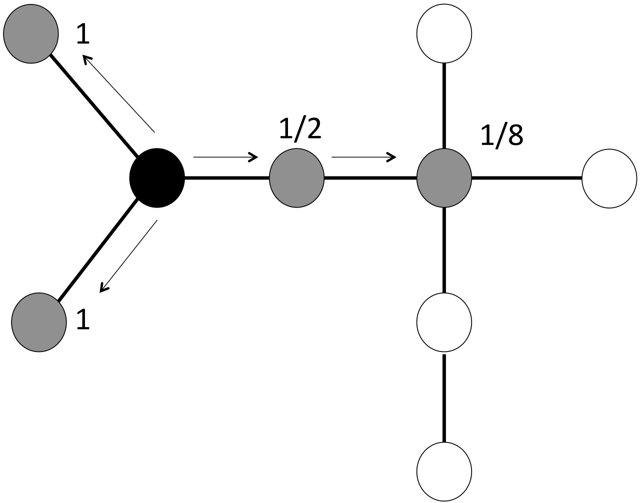
Corollary

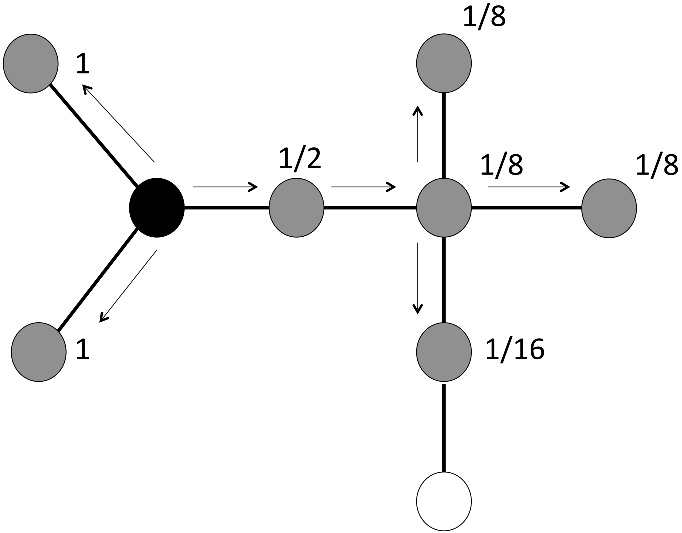
Suppose that G is a tree. Then, for all $i \in V$,

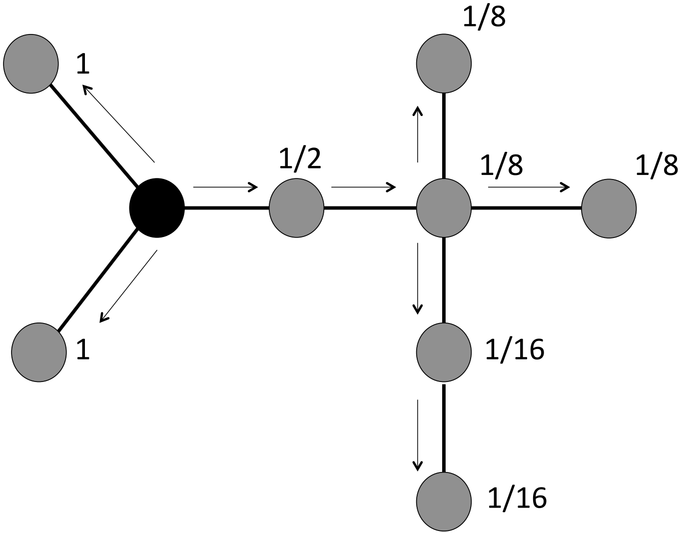
$$\mathcal{C}_i(G) = d_i(G) + 1.$$




1
1/2
1


1
1/2
1/8
1


1/8
1
1/2
1/8
1/8
1
1/16


1/8
1
1/2
1/8
1/8
1
1/16
1/16

Cycle

Corollary

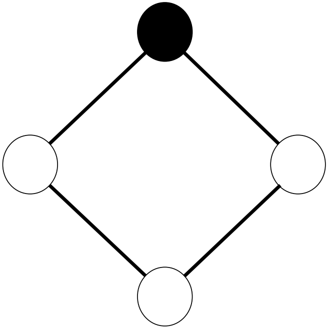
Suppose that G is a cycle of order n. Then, for all $i \in V$,

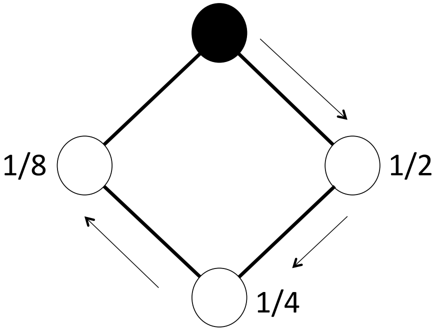
$$\mathcal{C}_i(G) = 3 - \frac{1}{2^{n-2}}$$

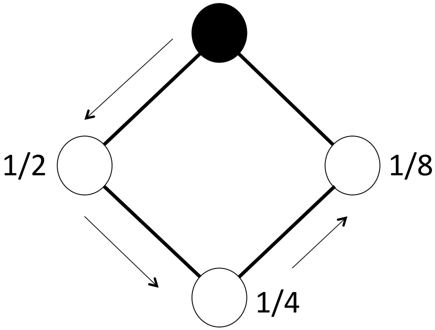
Proposition

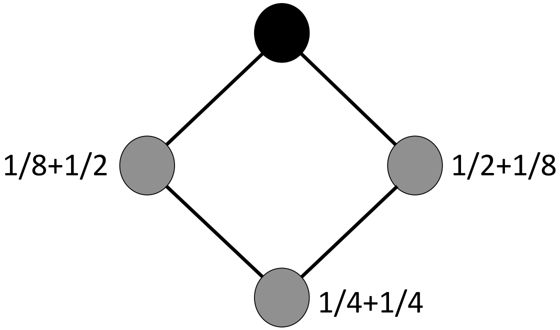
Consider a sequence of cycle graphs of order n, $\{G(n)\}_{n \in \mathbb{N}^+}$. Then, for all $i \in V$,

$$\lim_{n \to \infty} \mathcal{C}_i(G(n)) = 3$$




1/8
1/2
1/4


1/2
1/8
1/4


1/8+1/2
1/2+1/8
1/4+1/4

Complete graph

Corollary

Suppose that G is a complete graph of order n. Then, for all $i \in V$,

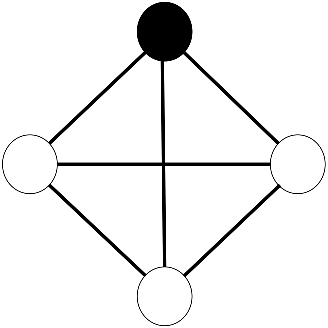
$$\mathcal{C}_i(G) = 1 + (n-1)\left(\sum_{i=1}^{n-1} \mathbf{P}(n-2, i-1)\left(\frac{1}{n-1}\right)^i\right)$$

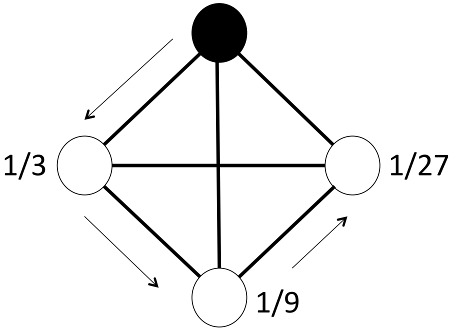
where $\mathbf{P}(n, i) \equiv \frac{n!}{(n-i)!}$ is number of ways of obtaining an ordered subset of i elements from a set of n elements.

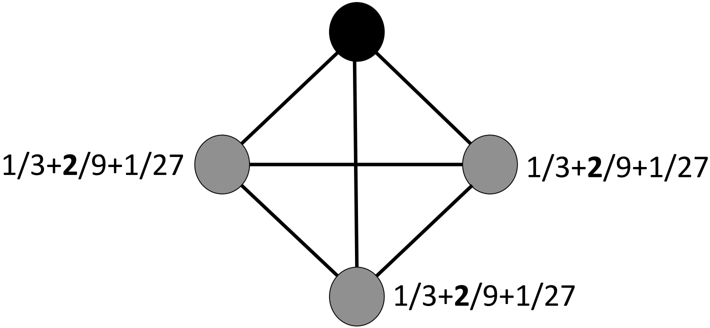
Proposition

Consider a sequence of complete graphs of order n, $\{G(n)\}_{n \in \mathbb{N}^+}$ Then, for all $i \in V$,

$$\lim_{n\to\infty} \frac{\mathcal{C}_i(G(n))}{\sqrt{n}} = \sqrt{\frac{\pi}{2}}$$




1/3
1/27
1/9


1/3+**2**/9+1/27
1/3+**2**/9+1/27
1/3+**2**/9+1/27

Random graphs

Proposition

Consider an Erdős-Rényi graph $G(n,\rho)$. Then, for a fixed n and a given $i \in V$, cascade centrality can be approximated by:

$$\mathcal{C}_i(G(n,\rho)) = 1 + (n-1)\rho - \frac{(n-1)(n-2)}{4}\rho^3 + o(\rho^4)$$

Conjecture

Consider an Erdős-Rényi graph $G(n,\rho)$. Then, for a fixed $\bar{d} = n\rho$:

$$\lim_{n\to\infty} \frac{\sum_{i\in N} \mathcal{C}_i(G(n,\rho))}{n} = \bar{d} + 1$$

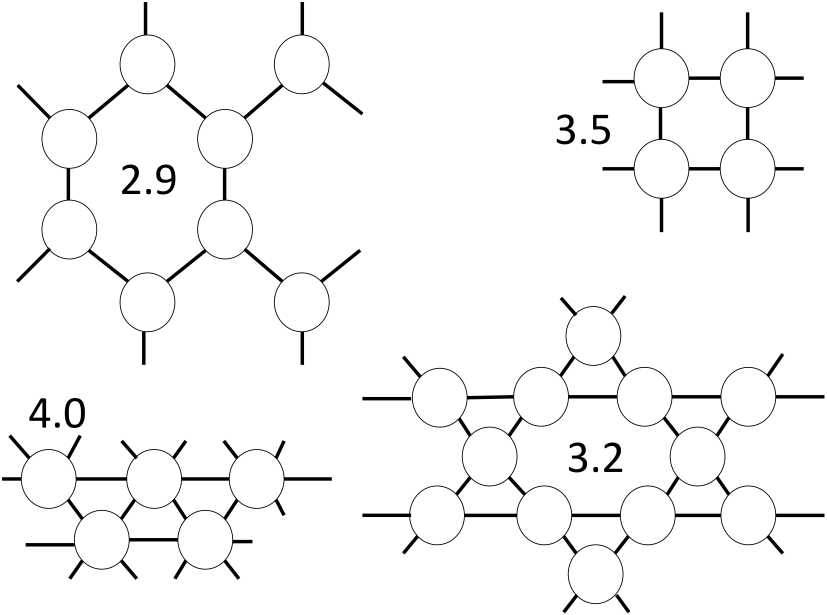
Lattices: self-avoiding walks

- For an approximation for cascade centrality in an infinite regular lattice, we can use the following proposition:

Proposition

Suppose that G is a r-regular, infinite lattice. Then, for all $i \in V$:

$$\mathcal{C}_i(G) \leq 1 + r$$


2.9
3.5
4.0
3.2

Conclusions

- Using a new notion of *cascade centrality*, we analyzed a tractable cascade process on general networks.
- We showed how these insights can help understand which networks prevent or help cascades.

Conclusions

- Using a new notion of *cascade centrality*, we analyzed a tractable cascade process on general networks.
- We showed how these insights can help understand which networks prevent or help cascades.

Future research questions

- In the next talks, I'll cover competition, pricing and network design: there will be plenty of research questions.
- In the meantime, cascade centrality for classes of random graphs (Erdős-Rényi or power-law) is an open problem.

Future research questions

- In the next talks, I'll cover competition, pricing and network design: there will be plenty of research questions.
- In the meantime, cascade centrality for classes of random graphs (Erdős-Rényi or power-law) is an open problem.

References I

D. Acemoglu, A. Ozdaglar, and M.E. Yildiz. Diffusion of innovations in social networks. In *Proc. of IEEE Conference on Decision and Control (CDC)*, 2011.

L. Blume, D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos. Which networks are least susceptible to cascading failures? In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, 2011.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *SIGKDD'03 Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

Stephen Morris. Contagion. *Review of Economic Studies*, 67(1):57–78, 2000.

H. Peyton Young. The diffusion of innovations in social networks. In *Economy as an Evolving Complex System*, volume 3, pages 267–282. Oxford University Press, 2006.