Approximating network reliability[1] and Birnbaum importance[2]

Stephen Eubank, Madhurima Nath, Yihui Ren

NSSAC-TR-19-005

SIAM Workshop on Network Science
Snowbird
May 23, 2019

[1]E.F. Moore and C.E. Shannon, Reliable circuits using less reliable relays, Journal of the Franklin Institute, 262, September, 1956.
[2]Z.W. Birnbaum, in Multivariate Analysis II, Proceedings of the 2nd International Symposium on Multivariate Analysis, ed. by P.R. Krishnaiah (Academic Press, New York, 1969), pp. 581–592

eubank@virginia.edu

Why Moore-Shannon reliability and Birnbaum importance matter

What's the most important edge in a network?

Why Moore-Shannon reliability and Birnbaum importance matter

~~What's the most important edge in a network?~~

- reliability – the probability that a dynamical system reaches a configuration with property $\mathcal{P}$.
- Birnbaum importance – the contribution of a set of interactions to the reliability.

What is the Birnbaum importance of a set of interactions with respect to a property $\mathcal{P}$ under dynamics $\mathcal{D}$?

Why do we need to approximate?

Birnbaum importance is the difference in reliability for 2 networks.

- Evaluating reliability is #P-hard.
- Interesting instances have $N \sim O(10^4 - 10^9)$ interactions.
- Birnbaum importance can be extremely small, e.g., 2^{-N}, so Monte Carlo may require exponentially many samples.
- There are 2^N possible sets of interactions to test.
- Birnbaum importance is not a submodular function, so greedy algorithms may not work.

eubank@virginia.edu

The thrilling conclusion – Don't Panic

For monotonic properties $\mathcal{P}$ under *SIR* dynamics – a class which includes cascading failures on layered networks – Birnbaum importance is well approximated by Bezier interpolation between weak- and strong-coupling approximations constructed from certain minimal sets of edges (cruxes) which can be sampled efficiently using Monte-Carlo. Moreover, Birnbaum importance is a submodular function of cruxes, so a greedy approach gives approximations with bounded errors.

Moore-Shannon network reliability

Definition: Reliability

Network reliability R is the probability that the dynamics D on an interaction network G generate a system configuration with property $\mathcal{P}$:

$$R \equiv \sum_{c \in \mathcal{C}} \mathcal{I}_{\mathcal{P}}(c) p_D(c) = \langle \mathcal{I}_{\mathcal{P}} \rangle = \sum_{\{c | \mathcal{I}_{\mathcal{P}}(c)=1\}} p_D(c)$$

where

- $\mathcal{C}$ is the set of all system configurations;
- $\mathcal{I}_{\mathcal{P}}(c)$ indicates whether configuration c has property $\mathcal{P}$;
- $p_D(c)$ is the probability that the dynamics generate configuration c.

A rose by any other name

$$R = \sum_{c \in \mathcal{C}} \mathcal{I}_{\mathcal{P}}(c) p_D(c) = \langle \mathcal{I}_{\mathcal{P}} \rangle$$

R is also known as

- Statistical Physics: a partition function;
- Computer Science: an instance of stochastic satisfiability;
- Graph Theory: an evaluation of the Tutte polynomial.
- Engineering: reliability

Example: Does S infect T under SIR dynamics on G?

$$R \equiv \sum_{c \in \mathcal{C}} \mathcal{I}_{\mathcal{P}}(c) p_D(c) = \langle \mathcal{I}_{\mathcal{P}} \rangle = \sum_{\{c | \mathcal{I}_{\mathcal{P}}(c) = 1\}} p_D(c)$$

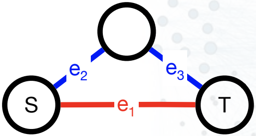
Definition: 1-1 mapping between events and edges in G

Event $E_{(i,j)}$ is an interaction between nodes i and j that leads to j being in the infected state, I. I.e., $E_{(i,j)} \equiv (i \text{ infects } j)$.

Remark: Dynamics as subgraph selection

- Configuration space $\mathcal{C}$ is the set of all subgraphs of G.
- $\mathcal{I}_{\mathcal{P}}(c)$ is true if and only if c contains a path from S to T.
- $p_D(c)$ is the product of all the transmission probabilities (edge weights) in the subgraph c.

$\mathcal{I}_{\mathcal{P}}$ is a deterministic binary function of N random binary inputs



Configurations that have property $\mathcal{P}$ (struts):

$E_1E_2E_3$	probability
TTT	$p_1p_2p_3$
TTF	$p_1p_2q_3$
TFT	$p_1q_2p_3$
TFF	$p_1q_2q_3$
FTT	$q_1p_2p_3$

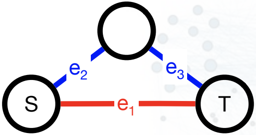
(with $q_i \equiv 1 - p_i$, $e_ie_j \equiv e_i \wedge e_j$)

SSAT: $p(e_1e_2e_3 \vee e_1e_2\overline{e}_3 \vee e_1\overline{e}_2e_3 \vee e_1\overline{e}_2\overline{e}_3 \vee \overline{e}_1e_2e_3)$

Reliability $= p_1p_2p_3 + p_1p_2q_3 + p_1q_2p_3 + p_1q_2q_3 + q_1p_2p_3$

eubank@virginia.edu

$\mathcal{I}_{\mathcal{P}}$ is a deterministic binary function of N random binary inputs



Configurations that have property $\mathcal{P}$ (struts):

$E_1 E_2 E_3$	probability
TTT	$p_1 p_2 p_3$
TTF	$p_1 p_2 q_3$
TFT	$p_1 q_2 p_3$
TFF	$p_1 q_2 q_3$
FTT	$q_1 p_2 p_3$

$E_1 E_2 E_3$	probability
T ##	p_1
#TT	$p_2 p_3$

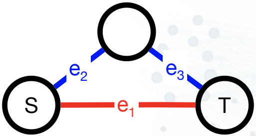
SSAT: $p(e_1 \vee e_2 e_3)$ $\leftarrow$ minimal struts $\in$ cruxes

$$\text{Reliability} = p_1 p_2 p_3 + p_1 p_2 q_3 + p_1 q_2 p_3 + p_1 q_2 q_3 + q_1 p_2 p_3$$
$$= p_1 + p_2 p_3 - p_1 p_2 p_3$$

eubank@virginia.edu

$\mathcal{I}_{\mathcal{P}}$ is a deterministic binary function of N random binary inputs

Configurations that have property $\mathcal{P}$ (struts):



$E_1E_2E_3$	probability
TTT	$p_1p_2p_3$
TTF	$p_1p_2q_3$
TFT	$p_1q_2p_3$
TFF	$p_1q_2q_3$
FTT	$q_1p_2p_3$

$E_1E_2E_3$	probability
T ##	p_1
#TT	p_2p_3

SSAT: $\quad p(e_1 \vee e_2e_3) \quad \leftarrow$ minimal struts $\in$ cruxes

$$\text{Reliability} = p_1p_2p_3 + p_1p_2q_3 + p_1q_2p_3 + p_1q_2q_3 + q_1p_2p_3$$
$$= p_1 + p_2p_3 \boxed{-p_1p_2p_3}$$

Corrects for counting the configuration TTT twice.

The inclusion-exclusion expansion corrects for overcounting

Definition: The inclusion-exclusion expansion

$$p(a_1 \vee a_2) = p(a_1) + p(a_2) - p(a_1 \wedge a_2)$$

Recursively,

$$p\Big(\bigvee_{i=1}^{M} a_i\Big) = \sum_{i=1}^{M} p(a_i) - \sum_{i>j=1}^{M} p(a_i \wedge a_j) + \sum_{i>j>k=1}^{M} p(a_i \wedge a_j \wedge a_k) - \cdots$$

For monotonic properties, this generates a Taylor series.

▸ monotonicity

Inclusion-exclusion for monotonic $\mathcal{P}$ generates a Taylor series

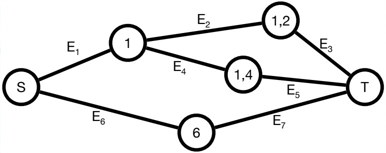
$$p(\bigvee_{i=1}^{M} a_i) = \sum_{i=1}^{M} p(a_i) - \sum_{i>j=1}^{M} p(a_i \wedge a_j) + \sum_{i>j>k=1}^{M} p(a_i \wedge a_j \wedge a_k) - \cdots$$

Example: $p(a_i) = x$ and a_i is independent of a_j

$$p(\bigvee_{i=1}^{M} a_i) = Mx - \binom{M}{2} x^2 + \binom{M}{3} x^3 - \cdots$$

eubank@virginia.edu

Example for S-T reliability under SIR dynamics



Remark: a more complicated case

a_1, a_2, and a_3 are compound events;
a_1 and a_2 are not independent.

$$\begin{aligned} R &= p_D\left((e_1 e_2 e_3) \vee (e_1 e_4 e_5) \vee (e_6 e_7)\right) \\ &\rightarrow x^2 + 2x^3 - 3x^5 + x^7 \end{aligned}$$

Note: $p(e_1 e_2 e_3 e_1 e_4 e_5) = p(e_1 e_2 e_3 e_4 e_5) \rightarrow x^5$.

Reliability has a natural dual

$$R = p\left(\bigvee_{i=1}^{M} a_i\right) = 1 - p\left(\bigwedge_{i=1}^{M} \neg a_i\right) = 1 - p\left(\bigvee_{j=1}^{\overline{M}} z_j\right),$$

a_i is a conjunction of independent events

$$a_i = \{e_{i,1}, e_{i,2}, \ldots, e_{i,m_i}\}\,; \quad p(a_i) = \prod_{k=1}^{m_i} p(e_{i,k})$$

z_j is a conjunction of independent "anti-events"

$$z_j = \{\overline{e}_{j,1}, \overline{e}_{j,2}, \ldots, \overline{e}_{j,\overline{m}_j}\}\,; \quad p(z_j) = \prod_{\ell=1}^{\overline{m}_j} [1 - p(e_{j,\ell})]$$

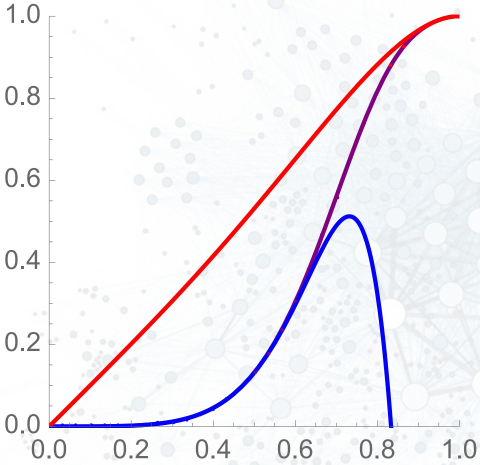
SSAT

The dual generates a strong-coupling expansion

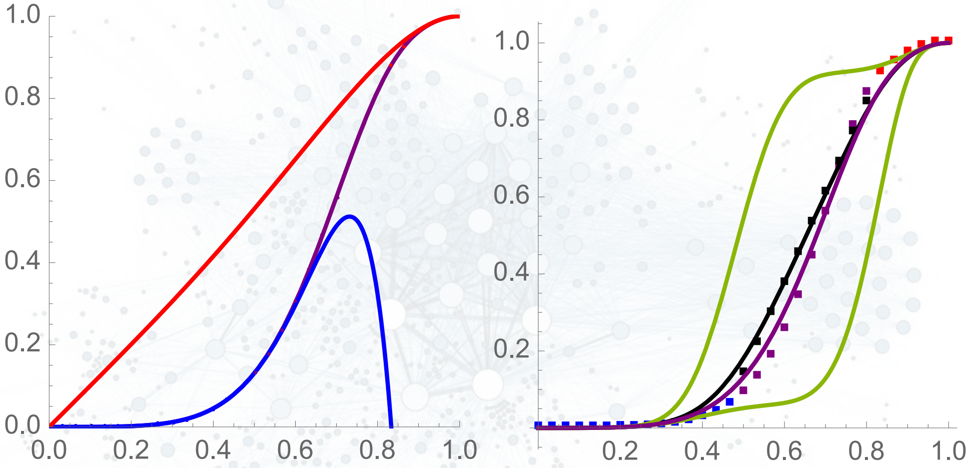
Remark: Taylor series as perturbative expansions

The probability of an interaction between two nodes, x, is a coupling constant in the dynamics. Truncated Taylor series for R at $x = 0$ are "perturbative estimates" for R, known in statistical physics as *weak-coupling* expansions.
Strong-coupling expansions are truncated Taylor series at the "other end" of the domain, here $x = 1$.

Weak / strong coupling expansions for $R(x)$ often diverge



Bezier approximations smoothly interpolate with tight bounds



Taylor series are expressed in power bases: x^k or $(1-x)^k$

Bezier interpolates in Bernstein basis: $\binom{N}{k} x^k (1-x)^{N-k}$

▸ Details

Leave-N-out change in reliability measures influence

Definition: Birnbaum importance

The Birnbaum importance of a subset of network elements $\mathcal{E}$ is the difference in reliability between the network with and without those elements.

$$B_{\mathcal{P}}(\mathcal{G}, \mathcal{E}) = R_{\mathcal{P}}(\mathcal{G}) - R_{\mathcal{P}}(\mathcal{G} \setminus \mathcal{E})$$

Remark: Birnbaum importance can be tiny

$\left[R_{\mathcal{P}}(\mathcal{G}) - R_{\mathcal{P}}(\mathcal{G} \setminus \mathcal{E})\right] / R_{\mathcal{P}}(\mathcal{G})$ can be $O(2^{-N})$.

Remark: Greedy approximation is possible

B is *not* submodular in edges, but it *is* submodular in cruxes.

eubank@virginia.edu

R is affine in each interaction probability

$$p(e_i \wedge e_i) = p(e_i) \equiv w_i \iff \frac{\partial R}{\partial w_i} = B(\{e_i\})/w_i$$

Remark: B defines R's sensitivity / tangent plane

B turns the inherently discrete edge ranking problem into a continuous gradient descent problem.

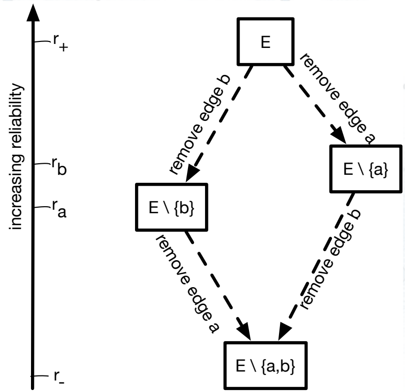
$$\frac{\partial R}{\partial w_i} = R|_{w_i=1} - R|_{w_i=0}$$

Remark: Recurrent contraction – deletion defines Tutte

$w_i = 1 \to$ contraction; $w_i = 0 \to$ deletion.

eubank@virginia.edu
University of Virginia
Biocomplexity Institute & Initiative

Improving numerical stability for SIR (and ??) dynamics



Evaluate $r_b - r_a$ directly, instead of r_b and r_a separately.

What you can do with these tools

- Compare networks ▸ addHealth
- Define effective couplings ▸ renormalization
- Graph reduction minimizing ΔR ▸ reduction
- Rank edges, find crossover in importance ▸ crossover
- Find communities in weighted, directed graphs ▸ example
- Generate graph duals ▸ duals

eubank@virginia.edu

(Almost automatic) extensions

- Reliability and Birnbaum importance for vertices
- Reliability and Birnbaum importance on hyper-graphs
- Other dynamics, especially time-dependent networks
- Bezier interpolation in the thermodynamic limit ($N \to \infty$)
- Combine MC and perturbative expansions
- Find optimal edge weight reductions, not removals
- Compare this notion of duality to others
- Detect graph isomorphisms ??

eubank@virginia.edu

The thrilling conclusion – Don't Panic

For monotonic properties $\mathcal{P}$ under *SIR* dynamics – a class which includes cascading failures on layered networks – Birnbaum importance is well approximated by Bezier interpolation between weak- and strong-coupling approximations constructed from certain minimal sets of edges (cruxes) which can be sampled efficiently using Monte-Carlo. Moreover, Birnbaum importance is a submodular function of cruxes, so a greedy approach gives approximations with bounded errors.

Nath, M., Ren, Y., Eubank, S., An approach to structural analysis using Moore-Shannon network reliability, in L.M. Aiello et al., Complex Networks 2018, pp. 537-549, 2019

Nath, M., Ren, Y., Khorramzadeh, Y., & Eubank, S. (2018). Determining whether a class of random graphs is consistent with an observed contact network. Journal of Theoretical Biology, 440, 121-132.

Ren, Y., Eubank, S., & Nath, M. (2016). From network reliability to the Ising model: A parallel scheme for estimating the joint density of states. Physical Review E, 94(4), 042125.

Khorramzadeh, Y., Youssef, M., Eubank, S., & Mowlaei, S. (2015). Analyzing network reliability using structural motifs. Physical Review E, 91(4), 042814.

Eubank, S., Youssef, M., & Khorramzadeh, Y. (2014). Using the network reliability polynomial to characterize and design networks. Journal of Complex Networks, 2(4), 356-372. doi:10.1093/comnet/cnu037.

Youssef, M., Khorramzadeh, Y., & Eubank, S. (2013). Network reliability: The effect of local network structure on diffusive processes. Physical Review E, 88(5), 052810.

Monotonicity, cruxes, cuts, and struts

Definition: Monotonicity

A property is *monotonic* if $I_{\mathcal{P}}(g) = 1 \Rightarrow I_{\mathcal{P}}(g') = 1 \, \forall g' \supseteq g$.

Monotonic properties admit 2 kinds of minimal sets we call "cruxes":

- cuts: Removing this set of interactions reduces the reliability to 0.
- struts: Guaranteeing that these interactions ensures the reliability is 1.

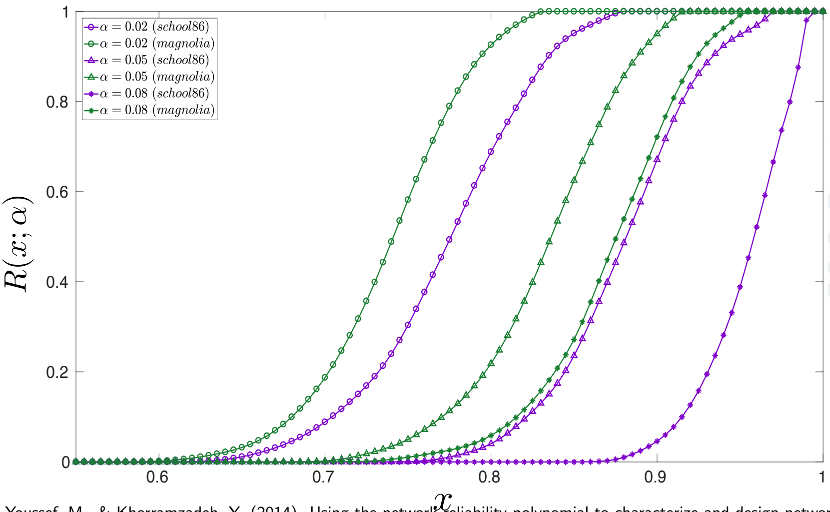
Variables in the SSAT instance induced by a monotonic property appear only in one sense, either e or $\bar{e}$, not both.

Cruxes can be found fairly quickly.

return details

Reliability can be used to compare networks

AddHealth's School 86 $\neq$ ERGM's Faux Magnolia!



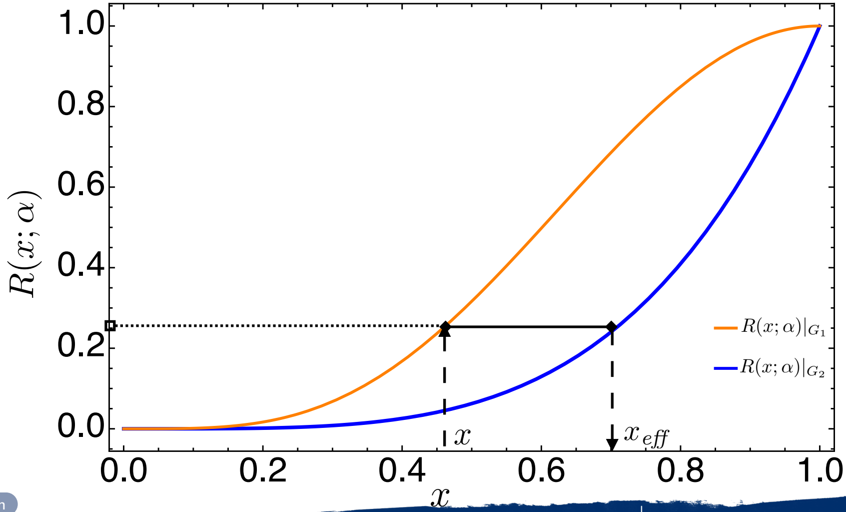
Eubank, S., Youssef, M., & Khorramzadeh, Y. (2014). Using the network reliability polynomial to characterize and design networks. Journal of Complex Networks, 2(4), 356-372. doi:10.1093/comnet/cnu037.

Nath, M., Ren, Y., Khorramzadeh, Y., & Eubank, S. (2018). Determining whether a class of random graphs is consistent with an observed contact network. Journal of Theoretical Biology, 440, 121-132.

◂ return ▸ details

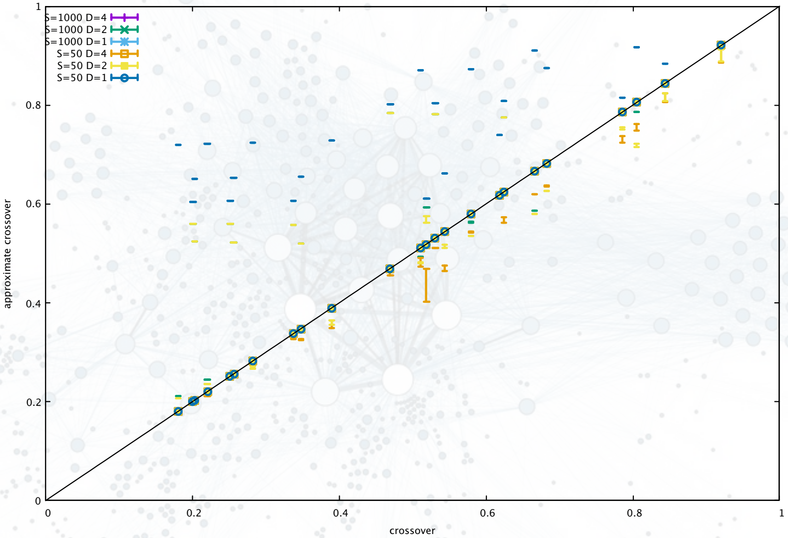
Renormalization can compensate for network differences

Define $x_{eff}(x)$ implictly: $R_{\mathcal{P}}(G_1, x) = R_{\mathcal{P}}(G_2, x_{eff})$



return

Ranking Edges: finding crossover in importance

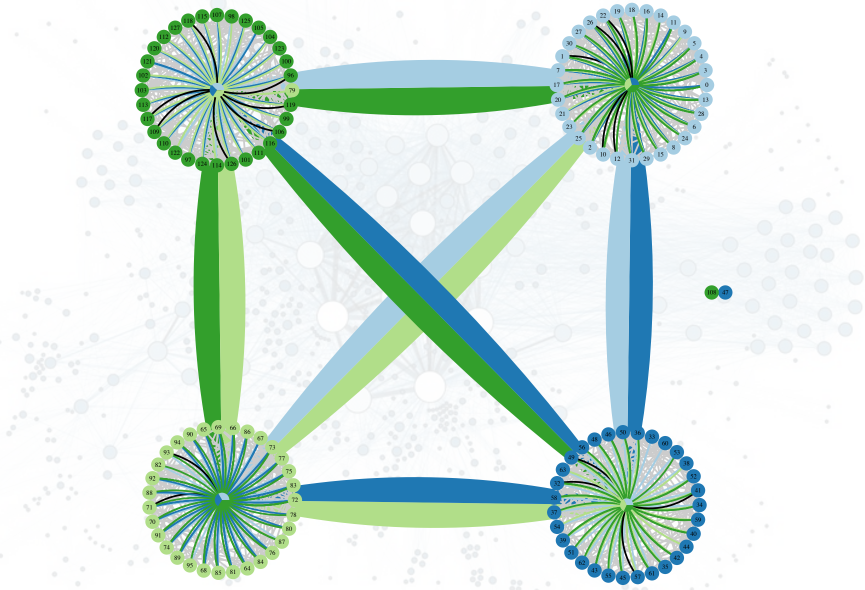


return

What is crossover?

eubank@virginia.edu

Detecting communities: finding planted partitions



◂ return

▸ details of the experiment

Graph reduction

Remark: Removing least important edges

By iteratively removing the k least important edges, we can construct the subgraph of size $E - k$ whose reliability is closest to the original graph's. Because of crossover effects, the reduced graph depends on dynamical parameters, e.g., x.

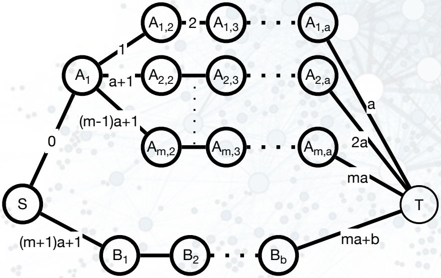
return

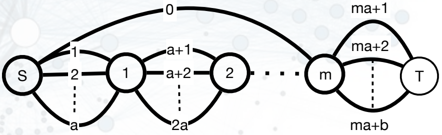
eubank@virginia.edu
University of Virginia Biocomplexity Institute & Initiative

Generating dual graphs

Example: Dual

The networks on the right are generated from those on the left by building a graph with a 1-1 correspondence between minimal struts in G and minimal cuts in G'.



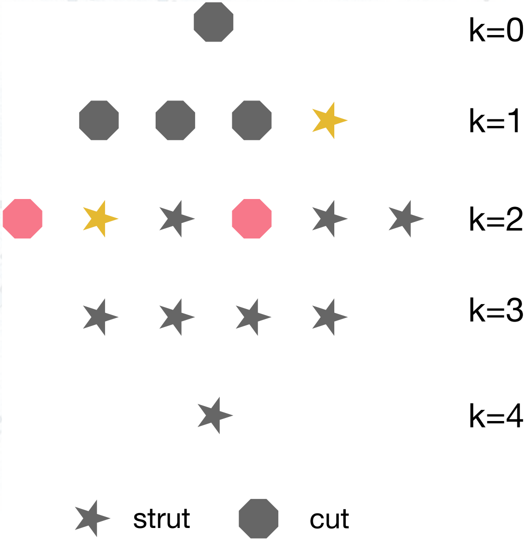


$$R_{G'}(x) = 1 - R_G(1 - x)$$

return

Finding cruxes

A lattice of the partially-ordered configurations

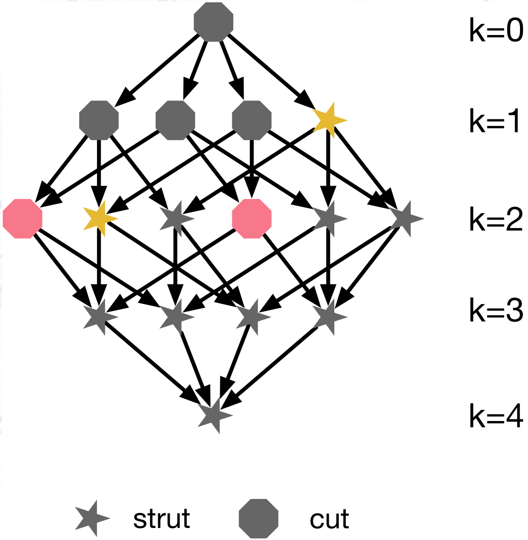


return

eubank@virginia.edu

Finding cruxes

A lattice of the partially-ordered configurations

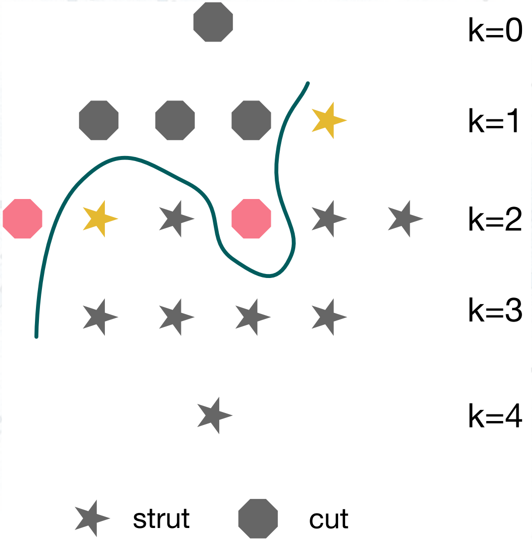


return

eubank@virginia.edu

Finding cruxes

Any path from top to bottom has one edge from a cut to a strut

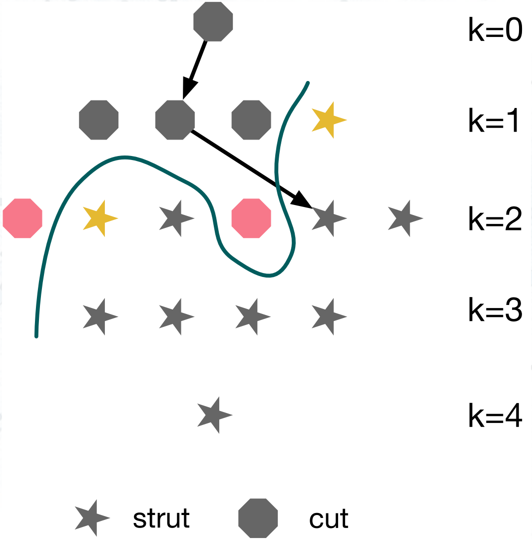


return

eubank@virginia.edu

Finding cruxes

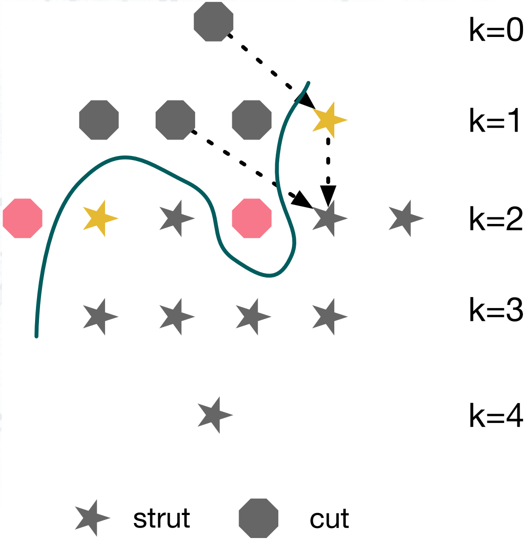
A simple search is guaranteed to lead to a minimal cut / strut



return

Finding cruxes

A simple search is guaranteed to lead to a minimal cut / strut



return

University of Virginia
Biocomplexity Institute & Initiative

Crossover

Remark: Crossover

The relative contribution of different edges to the overall reliability – the difference in their Birnbaum importance – depends on dynamical parameters, e.g., x. It is possible for the relative ranks to switch at certain values of x. For SIR dynamics on the family of toy networks, the ranks of the two most important edges switch when $x^{b-1} = 1 - (1 - x^{a-1})^m$.

return

eubank@virginia.edu

A community detection experiment

Remark: Community structure in a block stochastic or "planted ℓ-partition" graph

We iteratively removed the highest-rank edge from a 128-node, 1024-edge graph in which we had planted 4 communities. The average degree of the nodes is 16, of which 9 are within its planted community and 7 are external. The algorithm finds communities very similar to the planted ones across a wide range of parameters, and approximately 80% of the edges it removes are among the 448 planted inter-community edges.

return

eubank@virginia.edu

Bezier polynomials for two-point Taylor interpolation

$$f(x) = \sum_{i=0}^{N} \alpha_i x^i = \sum_{i=0}^{N} \overline{\alpha}_i (1-x)^i; \quad \hat{f}(x) = \sum_{k=0}^{N} \beta_k B(N,k,x),$$

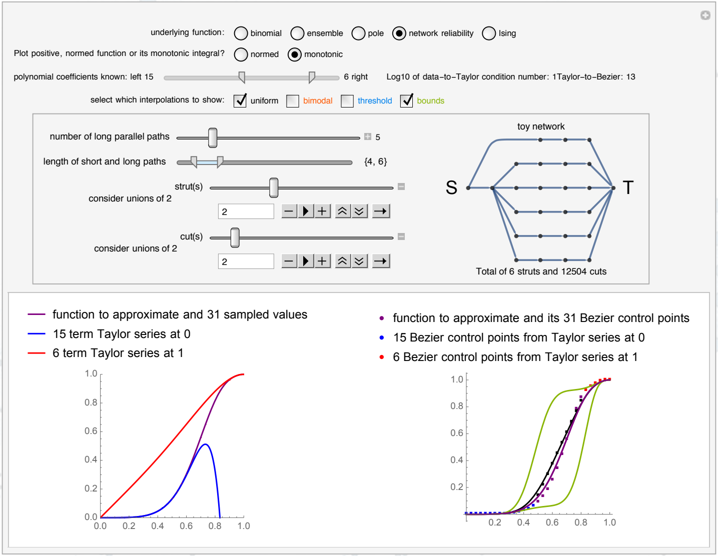
where $B(N,k,x) \equiv \binom{N}{k} x^k (1-x)^{N-k} = B(N, N-k, 1-x)$.

Given $\{\alpha_0, \ldots, \alpha_m\}$ and $\{\overline{\alpha}_0, \ldots, \overline{\alpha}_{\overline{m}}\}$, find $\{\beta_0, \ldots, \beta_m\}$ and $\{\beta_{N-\overline{m}}, \ldots, \beta_N\}$ such that the first m derivatives of f and $\hat{f}$ match at $x = 0$ and the first $\overline{m}$ derivatives of f and $\hat{f}$ match at $x = 1$.

$\hat{f}$ is a kernel density estimator for f.
For large N, the change of basis automatically produces an analytic continuation for f.

return

Coming soon: a Mathematica demo illustrating interpolation



return

eubank@virginia.edu

Details of the network comparison

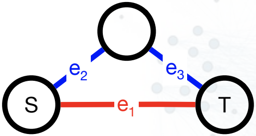
- AddHealth provides surveys of friendship networks in many schools.
- Faux Magnolia is a network available in the statnet R package that reproduces several local properties of the friendship network for one of the schools in AddHealth.
- The properties used to define reliability are that a fraction α of the population will be infected by a single, randomly selected infectious individual, for various α's.
- The results indicate that the graph structures responsible for the difference involve roughly 70-80% of the edges – local statistics cannot detect this difference.

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., Morris, M., 2016. ERGM: fit, Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (http://www.statnet.org). R package version 3.6.0.

return

$\mathcal{I}_{\mathcal{P}}$ is a deterministic binary function of N random binary inputs



Configurations that don't have property $\mathcal{P}$ (cuts):

$E_1E_2E_3$	probability
FTF	$q_1p_2q_3$
FFT	$q_1q_2p_3$
FFF	$q_1q_2q_3$

$E_1E_2E_3$	probability
FF#	q_1q_2
F#F	q_1q_3

SSAT: $p(\overline{e}_1\overline{e}_2 \vee \overline{e}_1\overline{e}_3)$ $\leftarrow$ minimal cuts $\in$ cruxes

$$\text{Reliability} = q_1p_2q_3 + q_1q_2p_3 + q_1q_2q_3$$
$$= q_1q_2 + q_1q_3 - q_1q_2q_3$$

return

eubank@virginia.edu