From Monte Carlo to Mountain Passes

Moments of Random Graphs With Fixed Degree Sequences

Phil Chodrow, MIT ORC

February 28th, 2020

Community Detection in Graphs

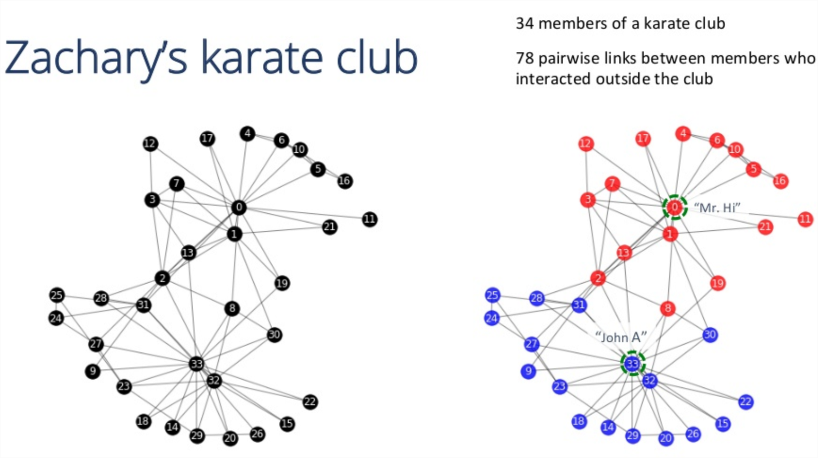


Figure from Erika Legara, "Community Detection with Networkx ." Link

Community Detection in Graphs

Ways to do community detection:

Inference: generative models

Dynamics: compression of random walks

Optimization: **modularity**, Min-Cut, Norm-Cut

A good review: Leto Peel, Daniel B Larremore, and Aaron Clauset. "The ground truth about metadata and community detection in networks". In: *Science Advances* 3.5 (2017), e1602548

The Modularity Objective

Let G be a non-loopy multigraph with adjacency matrix $\mathbf{W} \in \mathbb{Z}_+^n$.

Let $\mathbf{L} \in \{0,1\}^{n \times k}$ be a one-hot partitioning matrix into k labels.

The **modularity** of $\mathbf{L}$ is a number $Q(\mathbf{L}) \in [-1, 1]$ given by

$$Q(\mathbf{L}) = \frac{1}{\mathbf{e}^T \mathbf{W} \mathbf{e}} \mathrm{Tr}\left(\mathbf{L}^T \left[\mathbf{W} - \boldsymbol{\Omega} \right] \mathbf{L} \right)$$

$Q(\mathbf{L})$ is high when $\mathbf{L}$ assigns densely-connected pairs of nodes to the same label, and sparsely-connected pairs to different labels, **when compared to a null expectation $\boldsymbol{\Omega}$**.

Computing Ω

Usually, $\mathbf{\Omega} = \mathbb{E}_\eta[\mathbf{W}]$ is computed with respect to a *null* random graph η (a probability distribution over graphs). Which random graph?

Computing Ω

Usually, $\mathbf{\Omega} = \mathbb{E}_{\eta}[\mathbf{W}]$ is computed with respect to a *null* random graph η (a probability distribution over graphs). Which random graph?

The Physics Answer

Whichever random graph makes the expectation easy to compute. Stop bothering me.

Computing Ω

Usually, $\mathbf{\Omega} = \mathbb{E}_\eta[\mathbf{W}]$ is computed with respect to a *null* random graph η (a probability distribution over graphs). Which random graph?

The Physics Answer

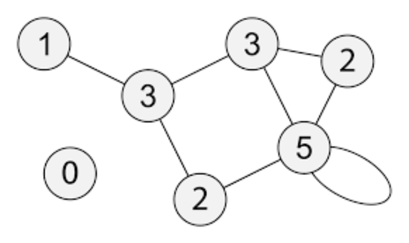
Whichever random graph makes the expectation easy to compute. Stop bothering me.

The Math Answer

The uniform distribution η over the space $\mathcal{G}_\mathbf{d}$ of non-loopy multigraphs with degree sequence $\mathbf{d}$.

Degree Sequence

The **degree** d_i of a node i is the number of edges incident to i.



The **degree sequence** contrains many of the macroscopic properties of a graph.[1]

[1] Mark E. J. Newman, S. H. Strogatz, and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: *Physical Review E* 64.2 (2001), p. 17.

Technical Goal

We want to:

Compute the expected adjacency matrix $\mathbb{E}_\eta[\mathbf{W}]$, where η is the uniform distribution on the set $\mathcal{G}_\mathbf{d}$ of multigraphs with degree sequence $\mathbf{d}$.

We want to:

Compute the expected adjacency matrix $\mathbb{E}_\eta[\mathbf{W}]$, where η is the uniform distribution on the set $\mathcal{G}_{\mathbf{d}}$ of multigraphs with degree sequence $\mathbf{d}$.

Problem

We don't know how to do this in practical time.

Agenda For Today

1. Introduce Markov Chain Monte Carlo for sampling from $\eta_{\mathbf{d}}$.
2. Derive/solve stationarity conditions on moments of $\eta_{\mathbf{d}}$.
3. Prove uniqueness of solution via a mountain-pass theorem.
4. Experiments.

A Note on My Working Process

So, I wrote this paper in, maybe, 2 months or so.

Then I submitted it so I could do job apps.

This had...consequences.

Moving on.

Markov Chain Monte Carlo

Main Idea

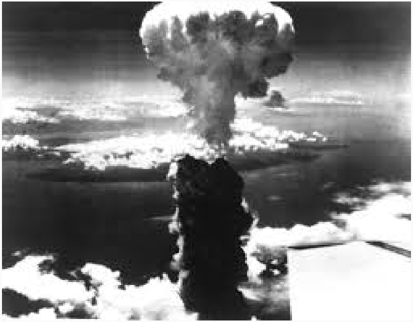
Sample from an intractable distribution μ by engineering a Markov chain whose stationary distribution is μ.

Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092

Markov Chain Monte Carlo

Main Idea

Sample from an intractable distribution μ by engineering a Markov chain whose stationary distribution is μ.



Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092

Example: 2d Gaussian

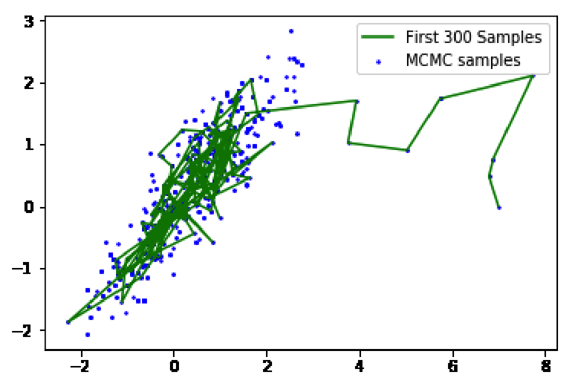


Image produced by Bernadita Ried Guachalla (University of Chile)

Edge-Swap MCMC

An **edge-swap** interchanges the endpoints of two edges, while preserving the degree sequence.

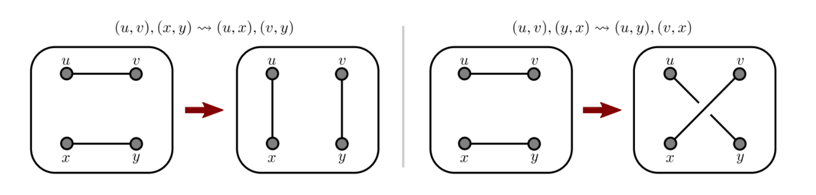
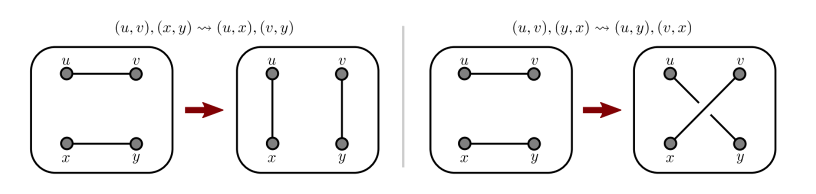


Image from Bailey K Fosdick et al. "Configuring random graph models with fixed degree sequences". In: *SIAM Review* 60.2 (2018), pp. 315–355

Edge-Swap MCMC

An **edge-swap** interchanges the endpoints of two edges, while preserving the degree sequence.



Theorem (Fosdick et al. 2018): We can do MCMC by *proposing* a random edge-swap on edges (i,j) and (k,ℓ) and *accepting* the swap with probability $w_{ij}^{-1} w_{k\ell}^{-1}$.

Image from Bailey K Fosdick et al. "Configuring random graph models with fixed degree sequences". In: *SIAM Review* 60.2 (2018), pp. 315–355

Markov Chain Monte Carlo for $\eta_{\mathbf{d}}$

Input: degree sequence $\mathbf{d}$, initial graph $G_0 \in \mathcal{G}_{\mathbf{d}}$, sample interval $\delta t \in \mathbb{Z}_+$, sample size $s \in \mathbb{Z}_+$.

Initialization: $t \leftarrow 0,\ G \leftarrow G_0$

for $t = 1, 2, \ldots, s(\delta t)$ **do**

  sample (i, j) and (k, ℓ) uniformly at random from $\binom{E_t}{2}$

  if $\mathit{Uniform}([0,1]) \leq \frac{1}{w_{ij} w_{k\ell}}$ **then**

    $G_t \leftarrow \mathrm{EdgeSwap}((i,j),(k,\ell))$

  else

    $G_t \leftarrow G_{t-1}$

Output: $\{G_t \text{ such that } t | \delta t\}$

Bailey K Fosdick et al. "Configuring random graph models with fixed degree sequences". In: *SIAM Review* 60.2 (2018), pp. 315–355

Stationarity Conditions

At stationarity of MCMC, we must have

$$\mathbb{E}_\eta[f(\mathbf{W}_{t+1}) - f(\mathbf{W}_t)] = 0$$

for all functions f.

Strategically picking f and handling a lot of algebra, we get the following theorems:

Low-Order Moments of $\eta_{\mathbf{d}}$

Theorem: There exists a vector $\boldsymbol{\beta} \in \mathbb{R}^n_+$ such that:

Indicators

$$\chi_{ij} \triangleq \eta_{\mathbf{d}}(w_{ij} \geq 1) \approx \frac{\beta_i \beta_j}{\mathbf{e}^T \boldsymbol{\beta}}$$

First Moments

$$\omega_{ij} \triangleq \mathbb{E}_{\eta}[w_{ij}] \approx \frac{\chi_{ij}}{1 - \chi_{ij}} \approx \frac{\beta_i \beta_j}{\mathbf{e}^T \boldsymbol{\beta} - \beta_i \beta_j}$$

We can provide precise (but fairly weak) error bounds on these approximations.

Computation of β

Since $\eta_{\mathbf{d}}$ is supported on graphs with degree sequence $\mathbf{d}$, we know that $\boldsymbol{\Omega}\mathbf{e} = \mathbf{d}$. Imposing this constraint, we get

$$h_i(\boldsymbol{\beta}) \triangleq \sum_{j \neq i} \frac{\beta_i \beta_j}{\mathbf{e}^T \boldsymbol{\beta} - \beta_i \beta_j} = d_i.$$

So, we can solve this to get $\boldsymbol{\beta}$. This is easy to do with standard iterative algorithms.

Me: “I did it! Publication please.”

Me: "I did it! Publication please."

Reviewer #1: "Prove uniqueness."

Me: "I did it! Publication please."

Reviewer #1: "Prove uniqueness."

Reviewer #2: "There are one thousand typos in this manuscript."

Me: "I did it! Publication please."

Reviewer #1: "Prove uniqueness."

Reviewer #2: "There are one thousand typos in this manuscript."

Me: ... :-(

Offscreen, Phil fixes one thousand typos.

A Month Later...

Theorem (Uniqueness of β)

Let

$$\mathcal{B} = \{\beta : \boldsymbol{\beta} \geq \mathbf{e}\ ,\ \max_i \beta_i^2 \leq \mathbf{e}^T \beta\}.$$

There exists at most one solution to the equation

$$h_i(\boldsymbol{\beta}) \triangleq \sum_{j \neq i} \frac{\beta_i \beta_j}{\mathbf{e}^T \boldsymbol{\beta} - \beta_i \beta_j} = d_i.$$

in $\mathcal{B}$.

Proof Outline

(a). The Jacobian of $\mathbf{h}$ is positive-definite on $\mathcal{B}$ (two pages of linear algebra tricks).

(b). The Hessian $\mathcal{H}(\boldsymbol{\beta})$ of the loss function $\mathcal{L}(\boldsymbol{\beta}) \triangleq \|\mathbf{h}(\boldsymbol{\beta}) - \mathbf{d}\|^2$ is positive-definite at all critical points of $\mathcal{L}$ (half a page more of linear algebra tricks)

> **Corollary:** all critical points of $\mathcal{L}$ are isolated local minima.

(c). **Mountain Pass Theorem**: $\mathcal{L}$ has at most one critical point.

Mountain Pass Theorem (Intuition)

If a “nice” function f has two, isolated local minima then f also has at least one more critical point which is **not** a local minimum.

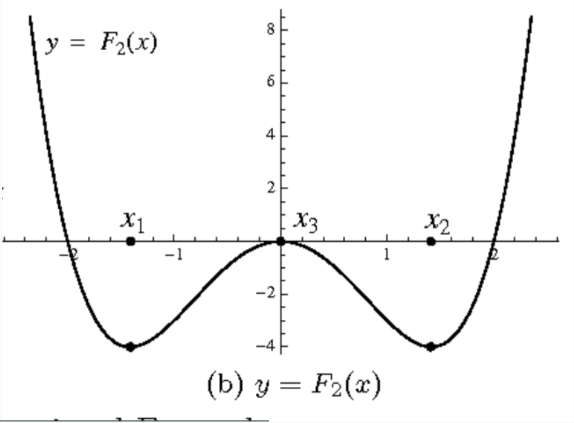


Figure from James Bisgard. “Mountain passes and saddle points”. In: *SIAM Review* 57.2 (2015), pp. 275–292

Mountain Pass Theorem (2-d)

The in multiple dimensions, the other critical point is usually a saddle point (the "mountain pass").

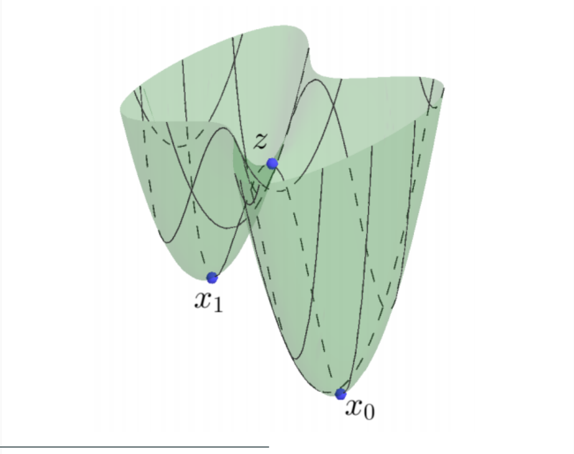


Figure from Lacey Johnson and Kevin Knudson. "Min-max theory for cell complexes". In: *arXiv:1811.00719* (2018)

Mountain Pass Theorem

Theorem (Mountain Pass Theorem in $\mathbb{R}^n$)

Suppose that a smooth function $q : \mathbb{R}^n \to \mathbb{R}$ satisfies the "Palais-Smale regularity condition." Suppose further that:

(a). $q(\mathbf{a}_0) = 0$.

(b). There exists an $r > 0$ and $\alpha > 0$ such that $q(\mathbf{a}) \geq \alpha$ for all $\mathbf{a}$ with $\|\mathbf{a} - \mathbf{a_0}\| = r$.

(c). There exists $\mathbf{a}'$ such that $\|\mathbf{a}' - \mathbf{a}_0\| > r$ and $q(\mathbf{a}') \leq 0$.

Then, q possesses a critical point $\tilde{\mathbf{a}}$ with $q(\tilde{\mathbf{a}}) \geq \alpha$.

James Bisgard. "Mountain passes and saddle points". In: *SIAM Review* 57.2 (2015), pp. 275–292, Antonio Ambrosetti and Paul H Rabinowitz. "Dual variational methods in critical point theory and applications". In: *Journal of Functional Analysis* 14.4 (1973), pp. 349–381

Proof Outline

$$h_i(\boldsymbol{\beta}) \triangleq \sum_{j \neq i} \frac{\beta_i \beta_j}{\mathbf{e}^T \boldsymbol{\beta} - \beta_i \beta_j} = d_i.$$

(a). The Jacobian of $\mathbf{h}$ is positive-definite on $\mathcal{B}$.

(b). The Hessian $\mathcal{H}(\boldsymbol{\beta})$ of the loss function $\mathcal{L}(\boldsymbol{\beta}) \triangleq \|\mathbf{h}(\boldsymbol{\beta}) - \mathbf{d}\|^2$ is positive-definite at all critical points of $\mathcal{L}$.

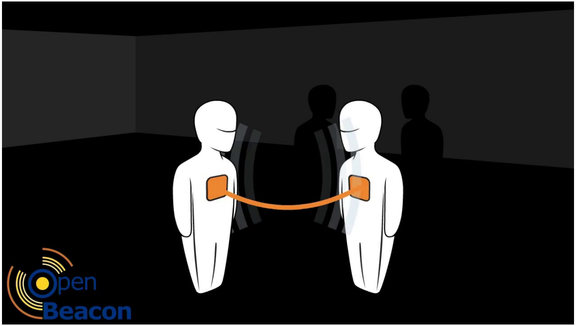
> **Corollary:** all critical points of $\mathcal{L}$ are isolated local minima.

(c). **Mountain pass theorem**: $\mathcal{L}$ has at most one critical point.

Ok, let's do some experiments.

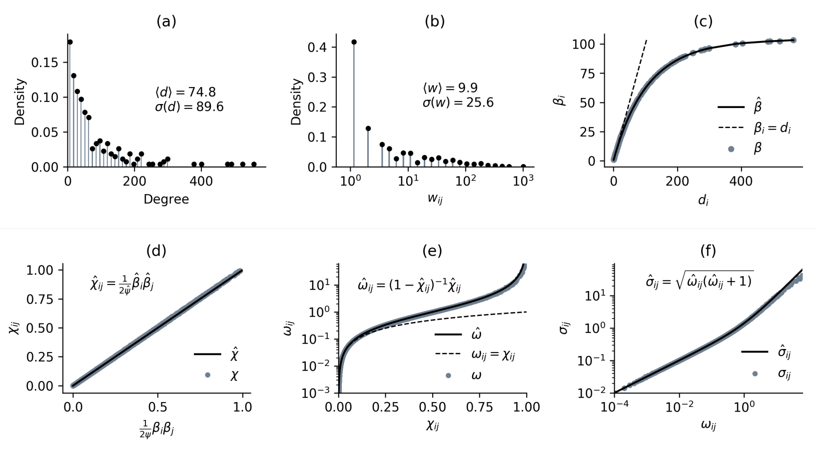
Data

Contact network in a French high school collected by the SocioPatterns project.[2]

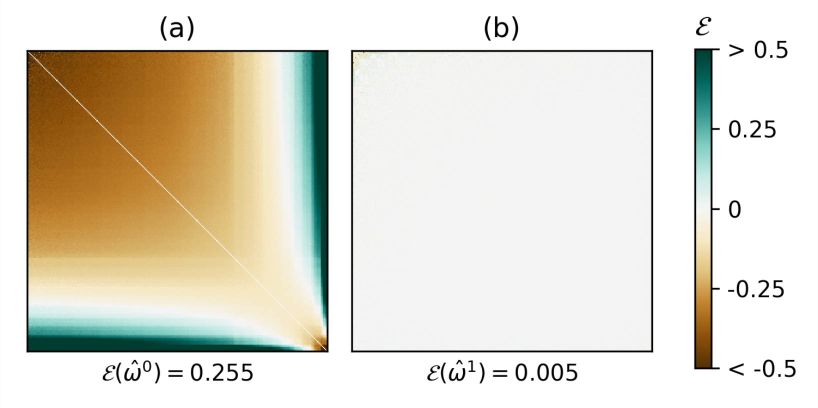


[2] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. "Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys". In: *PLOS ONE* 10.9 (2015). Ed. by Cecile Viboud, Austin R. Benson et al. "Simplicial closure and higher-order link prediction". In: *Proceedings of the National Academy of Sciences* 115.48 (2018), pp. 11221–11230.

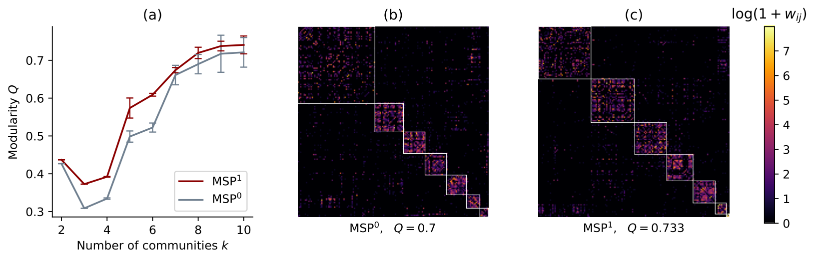
Numerical Test: High School Contact Network



Numerical Test: High School Contact Network



Modularity Maximization with the Uniform Null



Takeaways from This Work

(a) Both reviewers were ultimately very helpful (paper resubmitted).

Takeaways from This Work

(a) Both reviewers were ultimately very helpful (paper resubmitted).

(b) Uniform distributions are hard.

Takeaways from This Work

(a) Both reviewers were ultimately very helpful (paper resubmitted).
(b) Uniform distributions are hard.
(c) Surveilling high schoolers is fun (but only in the name of science).

Takeaways from This Work

(a) Both reviewers were ultimately very helpful (paper resubmitted).

(b) Uniform distributions are hard.

(c) Surveilling high schoolers is fun (but only in the name of science).

(d) Maybe don't write and submit papers in two months?...

Thanks!

arXiv.org > cs > arXiv:1909.09037

Computer Science > Social and Information Networks

Moments of Uniform Random Multigraphs with Fixed Degree Sequences

Philip S. Chodrow

(Submitted on 19 Sep 2019)

We study the expected adjacency matrix of a uniformly random multigraph with fixed degree sequence $\mathbf{d}$. This matrix arises in a variety
spreading processes. Its structure is well-understood for large, sparse, simple graphs: the expected number of edges between nodes i
the approximation no longer applies. We derive a novel estimator using a dynamical approach: the estimator emerges from the stationa
error bounds are available under mild assumptions, and the estimator can be computed efficiently. We test the estimator on a small net
standard expression. We then compare modularity maximization techniques using both the standard and novel estimator, finding that t
importance of using carefully specified random graph models in data scientific applications.

philchodrow.com

github.com/philchodrow

@philchodrow