Indiana University
Network Science Institute

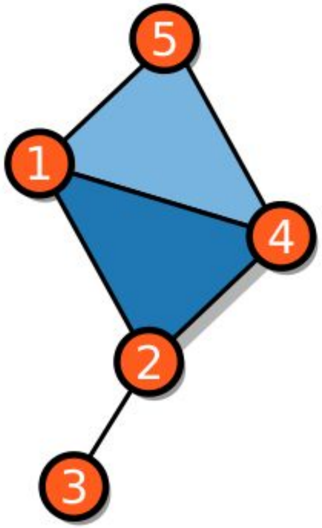
The simplicial configuration model

A Sampling Method

Alice Patania, Network Science Institute, Indiana University

A simplicial complex

A **simplicial complex** is a collection of simplices (non-empty finite sets ordered under inclusion).
It can be uniquely identified by its maximal simplices under inclusion (**facets**)



A simplicial complex

A **simplicial complex** is a collection of simplices (non-empty finite sets ordered under inclusion).
It can be uniquely identified by its maximal simplices under inclusion (**facets**)

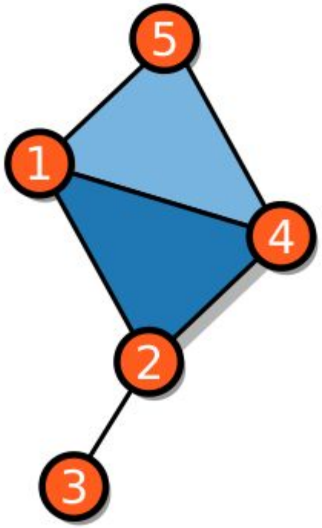
To every simplicial complex we can assign two sequences:

Degree sequence

Number of maximal simplices under inclusion incident on a node

Size sequence

Number of nodes contained in a maximal simplex



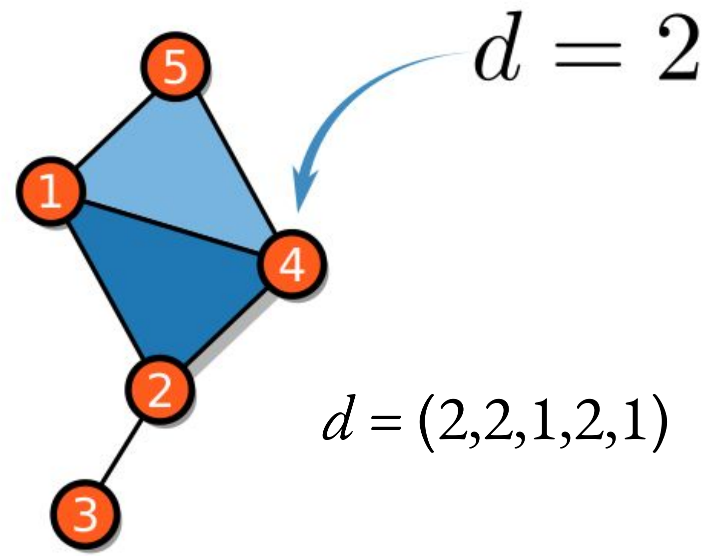
Simplicial degree and Facet size

A **simplicial complex** is a collection of simplices (non-empty finite sets ordered under inclusion).
It can be uniquely identified by its maximal simplices under inclusion (**facets**)

To every simplicial complex we can assign two sequences:

Degree sequence
Number of maximal simplices under inclusion incident on a node

Size sequence
Number of nodes contained in a maximal simplex



Simplicial degree and Facet size

A **simplicial complex** is a collection of simplices (non-empty finite sets ordered under inclusion).
It can be uniquely identified by its maximal simplices under inclusion (**facets**)

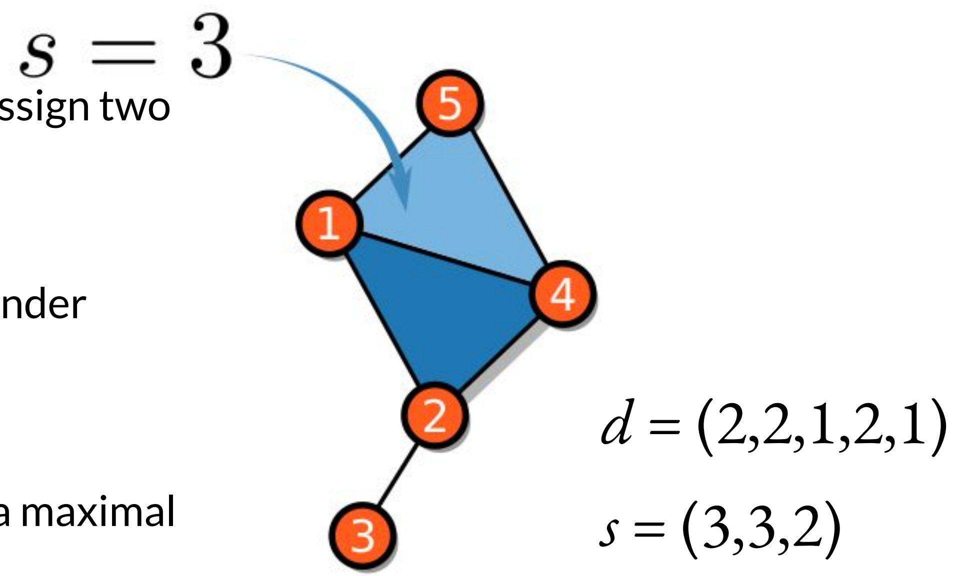
To every simplicial complex we can assign two sequences:

Degree sequence

Number of maximal simplices under inclusion incident on a node

Size sequence

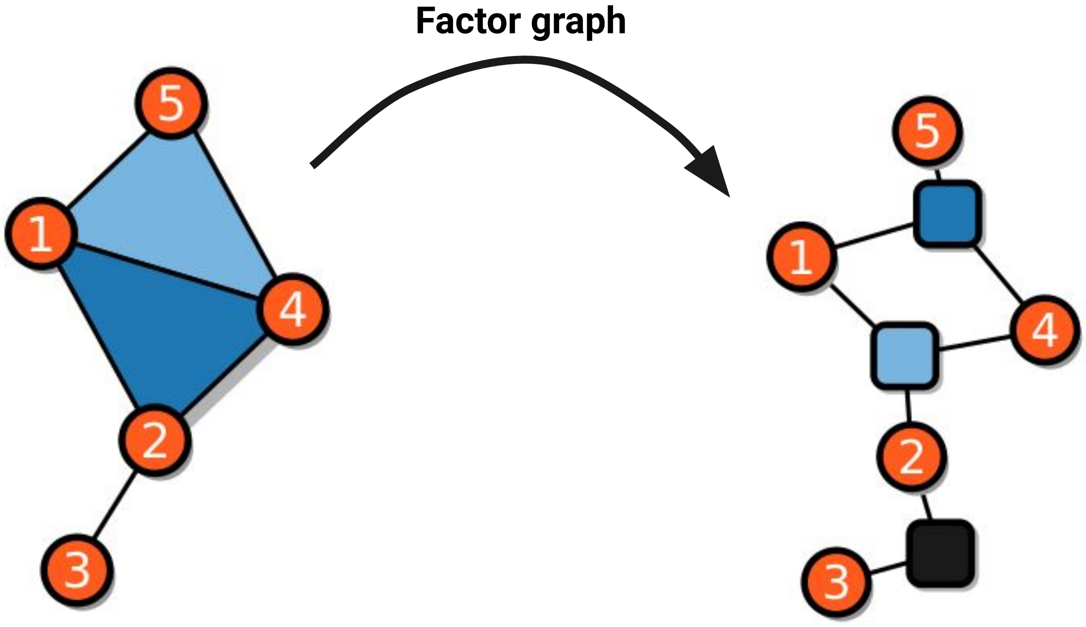
Number of nodes contained in a maximal simplex



Bipartite graphs and Simplicial complexes

Theorem

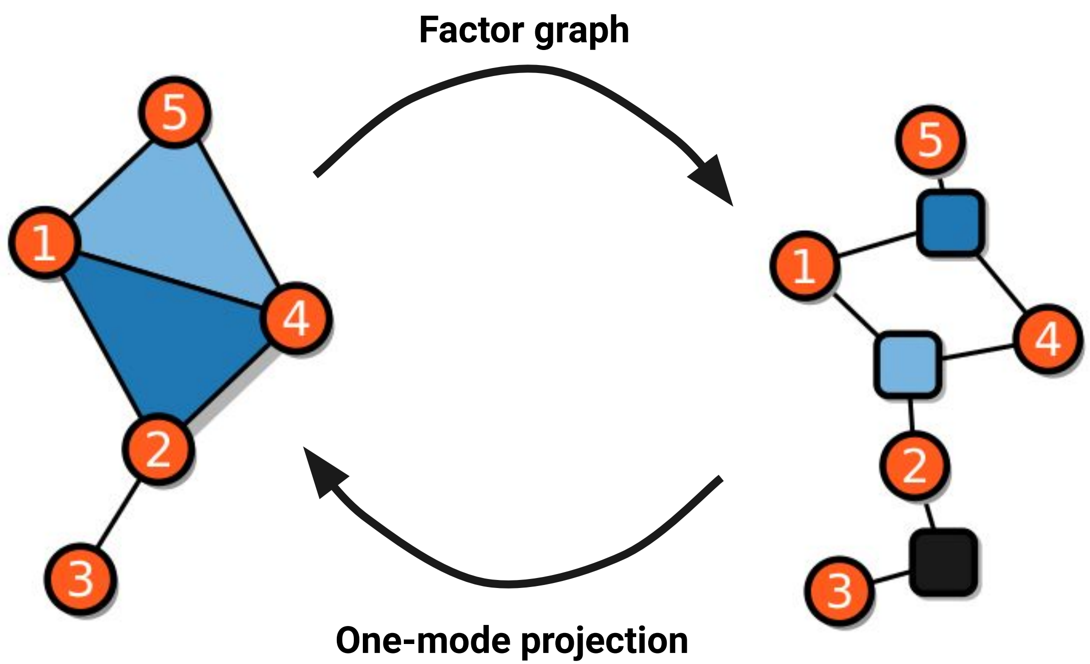
For every Σ, $\exists G$, bipartite graph, s.t. one of its two one-mode projections G_V is the underlying graph of Σ.
Moreover, the facet size sequence of Σ is equal to the degree sequence of F.



Bipartite graphs and Simplicial complexes

Theorem

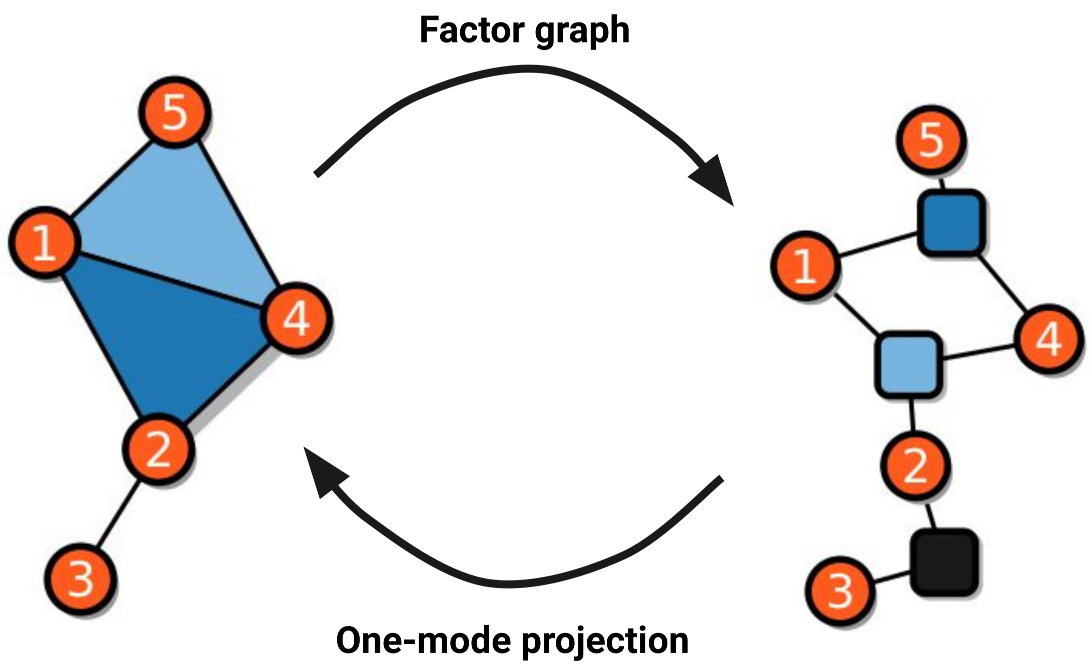
Let G be a bipartite graph with vertex sets $\{F, V\}$, G_V its one-mode projections onto the vertex set V.
Then it exists a simplicial complex Σ whose underlying graph is G_V.



Bipartite graphs and Simplicial complexes

Idea

Use existing random models for bipartite graphs to construct a sampling method for SCM.



Existing approaches

*this list is not exhaustive

Erdos-Renyi inspired:

Random pure simplicial complexes [Linial-Meshulam (2006)]
Random simplicial complexes [Kahle (2009)]
Multi-parameter random simplicial complexes [Costa-Farber (2015)]

Exponential random graphs inspired:

Exponential random simplicial complex [Zuev et al. (2015)]

Preferential attachment for simplicial complexes:

Network geometry with flavor [Bianconi-Rahmede (2016)]

Configuration model:

Configuration model for pure simplicial complexes [Courtney-Bianconi (2016)]

Configuration model

Idea
Use existing random models for bipartite graphs to construct a sampling method for SCM.

The **configuration model** is a generative model for random graphs with fixed number of nodes and degree sequence.

Configuration model

Idea
Use existing random models for bipartite graphs to construct a sampling method for SCM.

The **configuration model** is a generative model for random graphs with fixed number of nodes and degree sequence.

This implies the following are fixed:

number of vertices (0-simplices) and number of maximal simplices
vertices' degree sequence and maximal simplices' size sequence

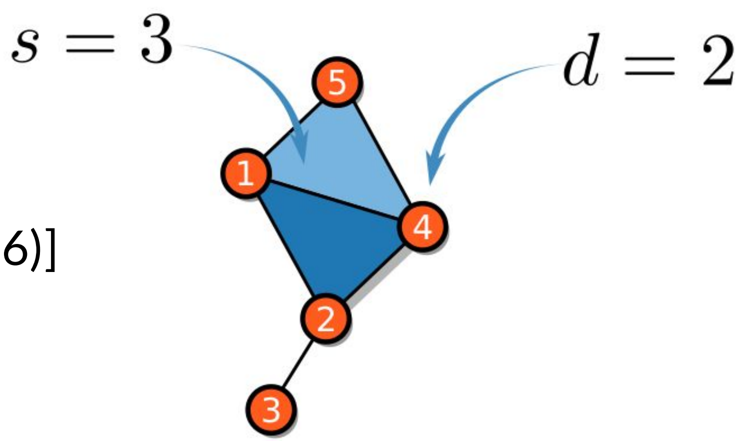
Simplicial degree and Facet size

The **simplicial configuration model** (SCM) is the distribution

$$\Pr(K; \boldsymbol{d}, \boldsymbol{s}) = 1/|\Omega(\boldsymbol{d}, \boldsymbol{s})|$$

Where $\Omega(\boldsymbol{d}, \boldsymbol{s})$: number of simplicial complexes with sequences $(\boldsymbol{d}, \boldsymbol{s})$

This approach generalizes a method by [Courtney and Bianconi, Phys. Rev. E 93, (2016)]

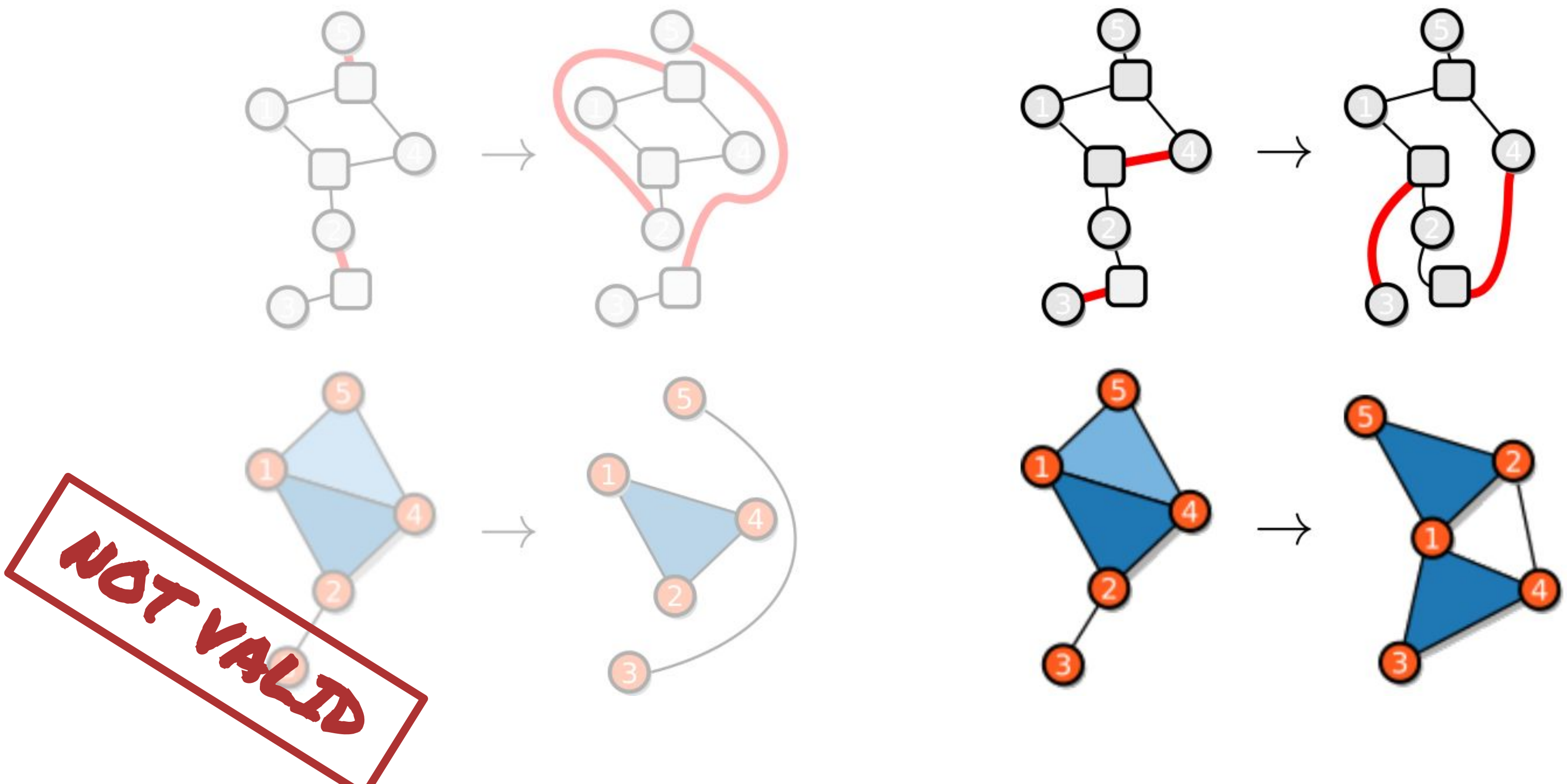


Bipartite graphs and Simplicial complexes

The **simplicial configuration model** (SCM) is the distribution

$$\Pr(K; \boldsymbol{d}, \boldsymbol{s}) = 1/|\Omega(\boldsymbol{d}, \boldsymbol{s})|$$

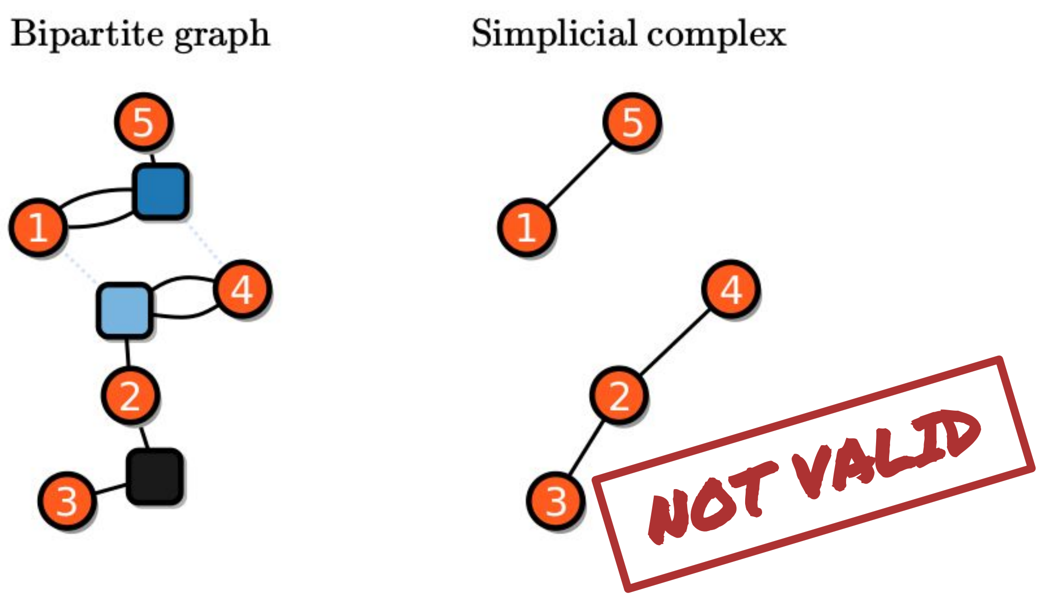
Where $\Omega(\boldsymbol{d}, \boldsymbol{s})$: number of simplicial complexes with sequences $(\boldsymbol{d}, \boldsymbol{s})$



Adding constraints

First constraint: **No multi-edges**

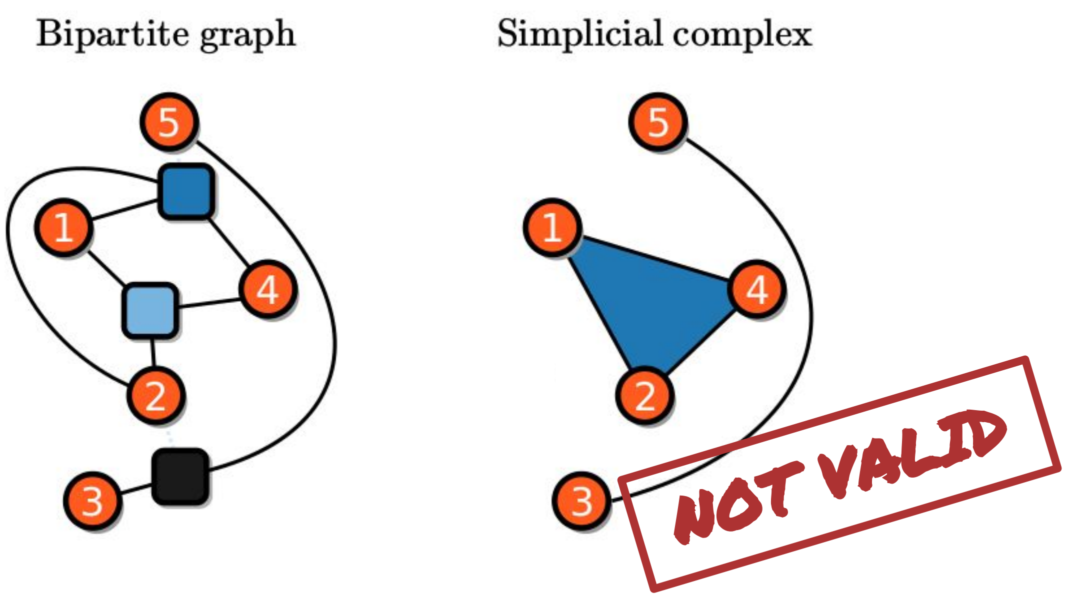
Multi-edges decrease the size of the maximal simplices.



Adding constraints

Second constraint: **No included neighborhoods**

Included neighborhoods violate the maximality assumption of the facets.



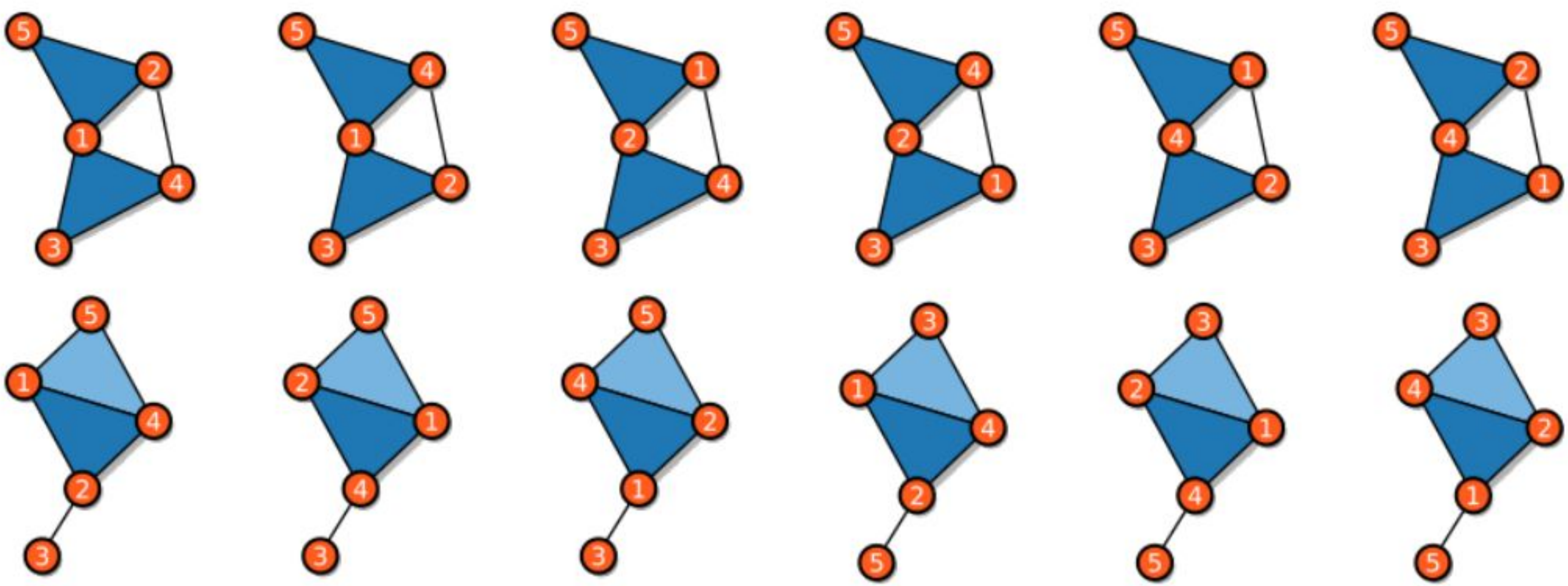
Adding constraints

Constraints:

No multi-edges

No included neighborhoods

Then the acceptable configurations for the toy example are the following:

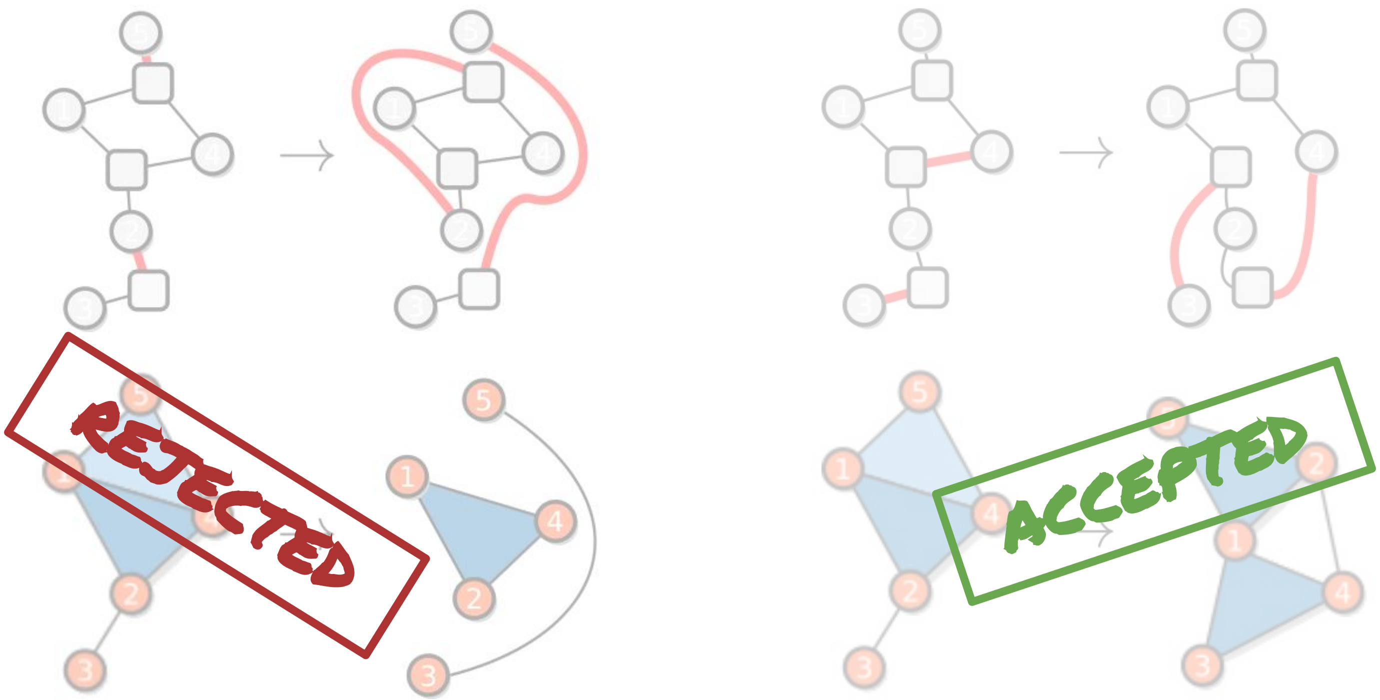


Bipartite graphs and Simplicial complexes

The **simplicial configuration model** (SCM) is the distribution

$$\Pr(K; \boldsymbol{d}, \boldsymbol{s}) = 1/|\Omega(\boldsymbol{d}, \boldsymbol{s})|$$

Where $\Omega(\boldsymbol{d}, \boldsymbol{s})$: number of simplicial complexes with sequences $(\boldsymbol{d}, \boldsymbol{s})$

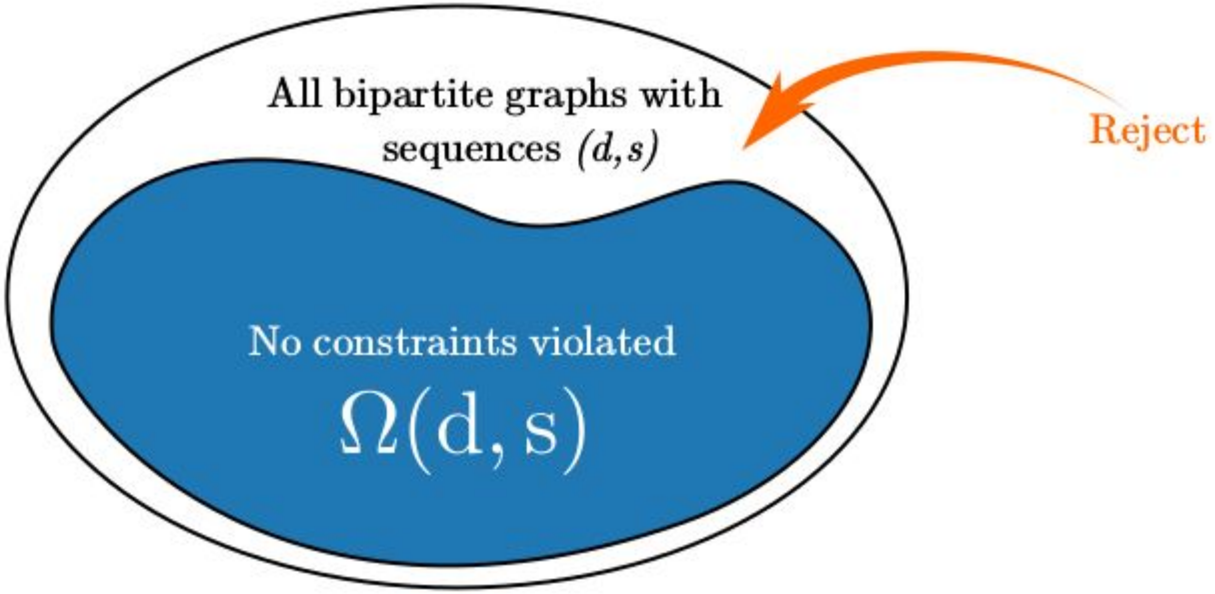


Adding constraints

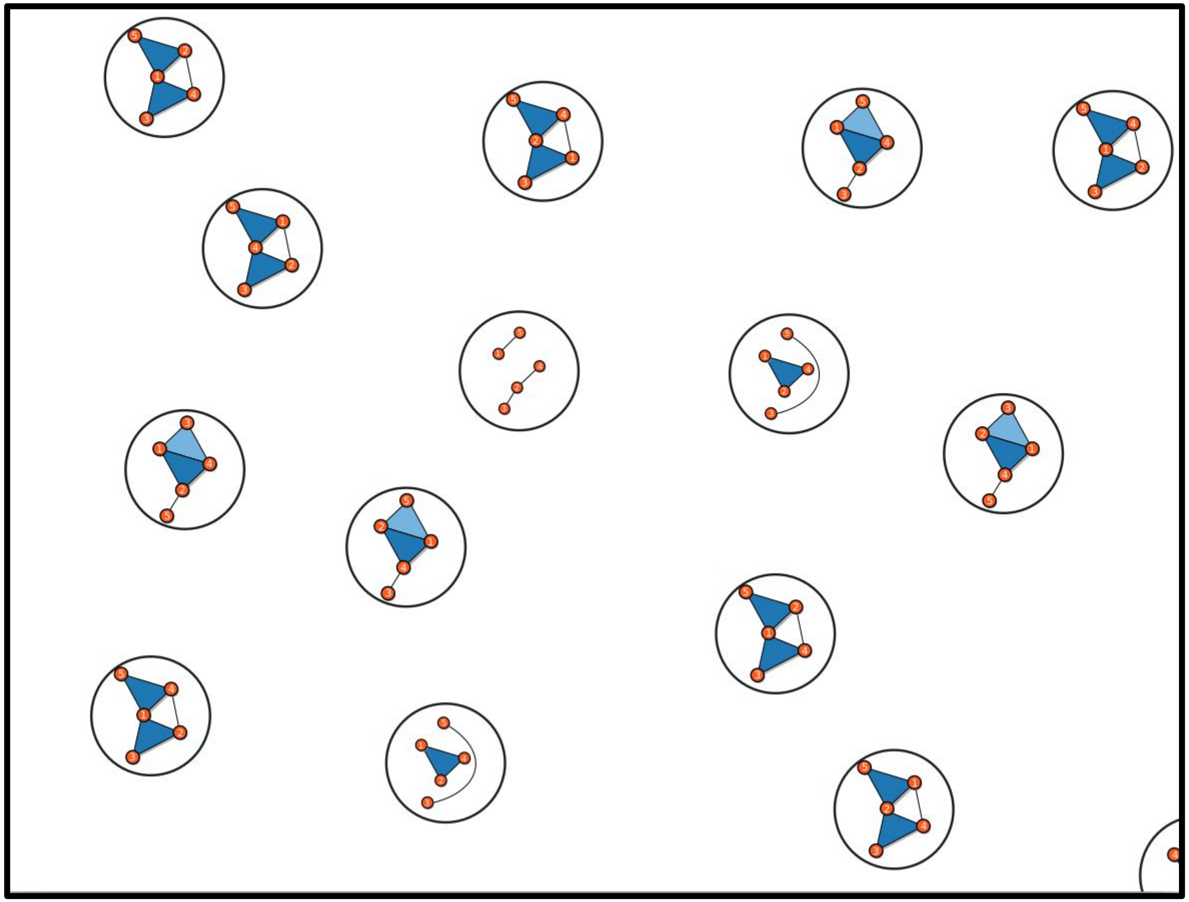
! **Problem with rejection sampling**: Far too many rejections!

Loose upper bound :

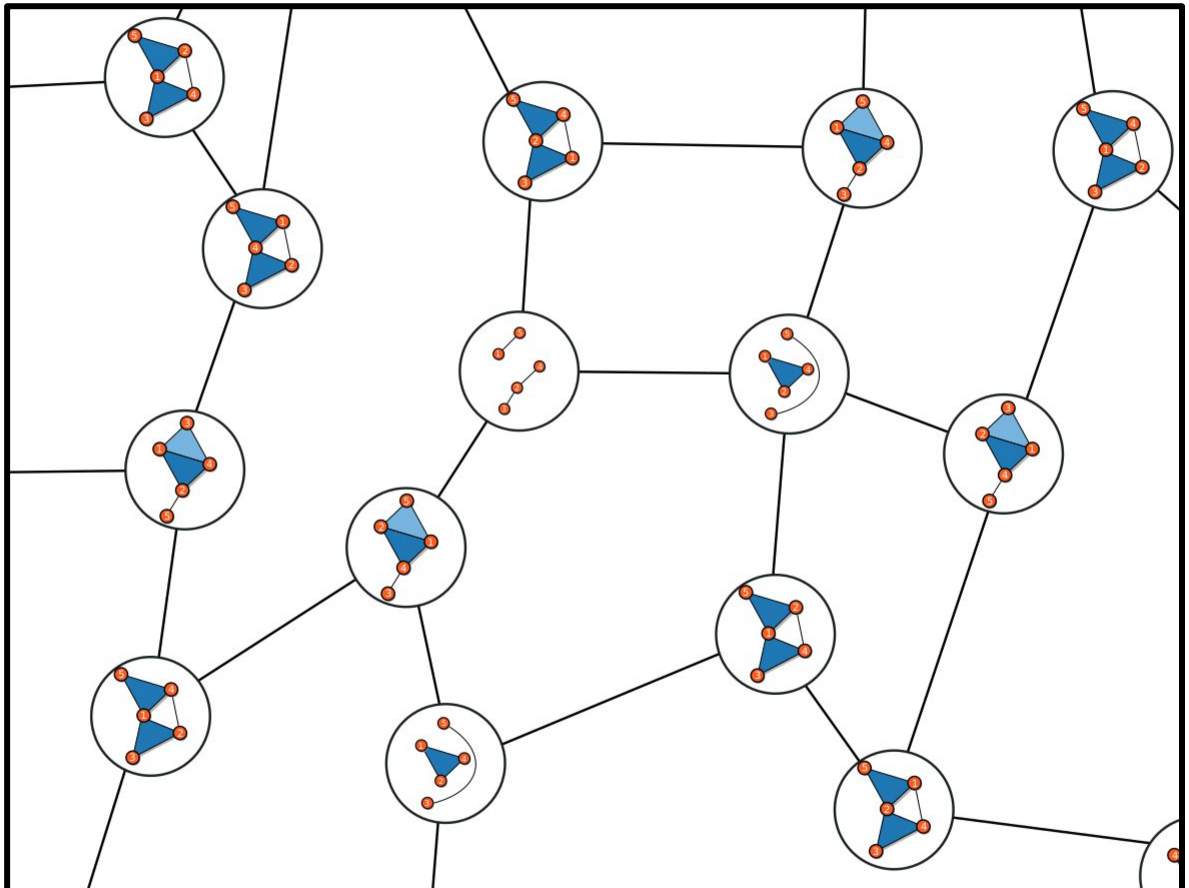
$$\Pr[\text{reject}] > \exp[-0.5(\langle d^2\rangle/\langle d\rangle - 1)(\langle s^2\rangle/\langle s\rangle - 1)]$$



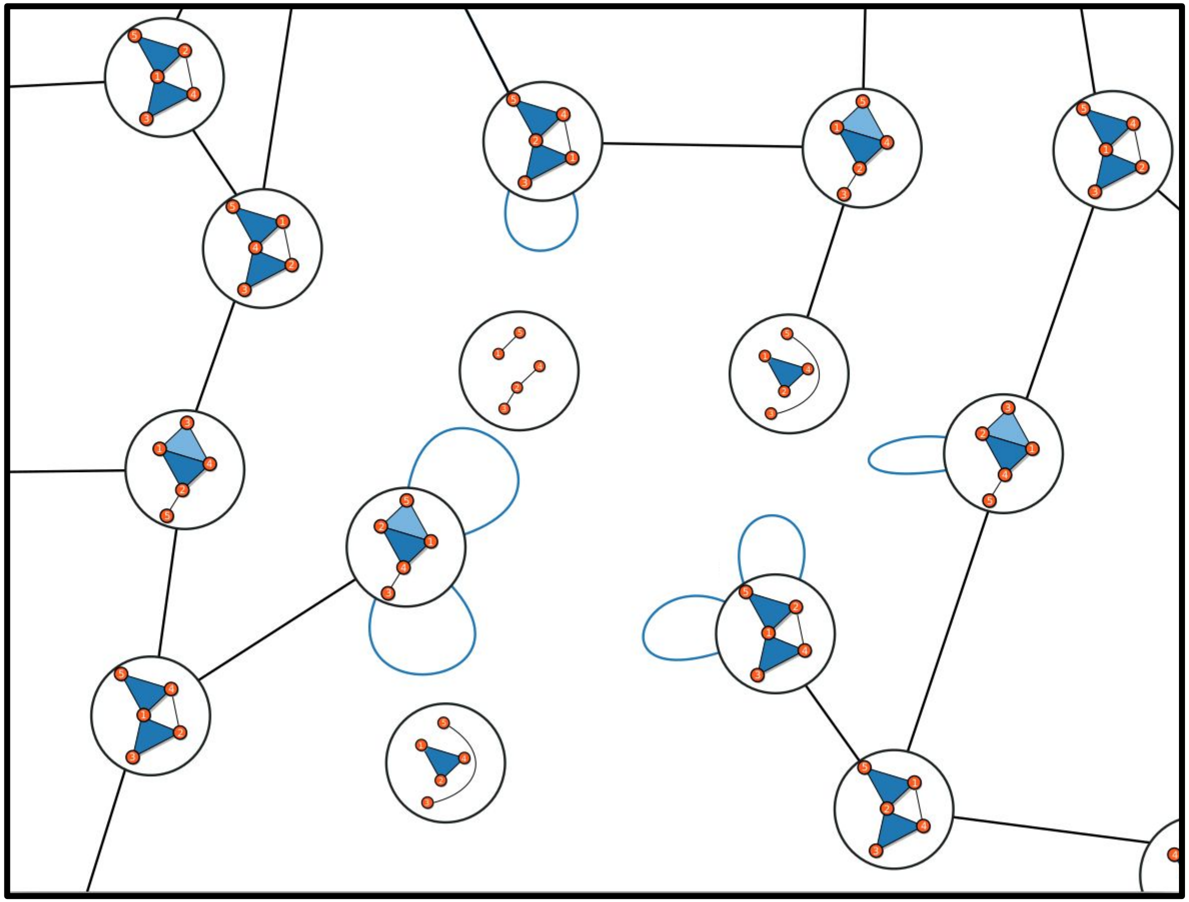
Markov Chain Monte Carlo sampling



Markov Chain Monte Carlo sampling



Markov Chain Monte Carlo sampling



MCMC sampling: The details

Move set

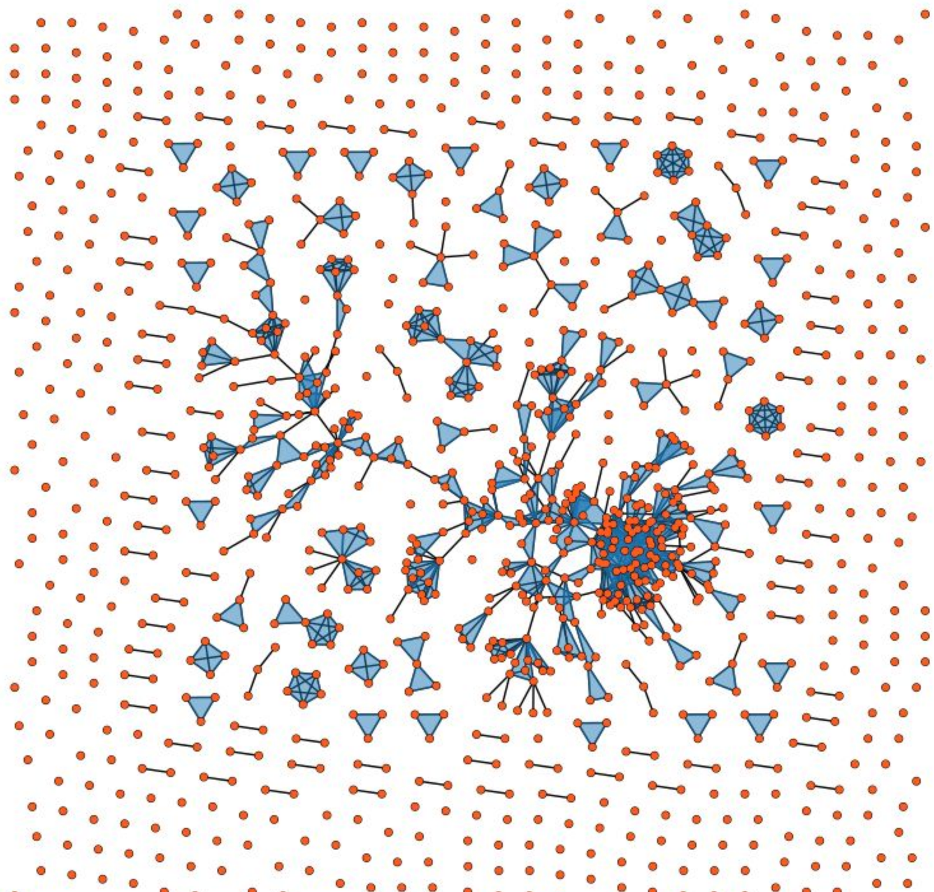
1. Pick $L \sim P$ random edges in bipartite graph
   P can be arbitrary, we use $\Pr[L = \ell] = \exp[\lambda \ell]/Z$
2. Rewire edges. If multi-edge or included neighbors, reject.

Similar to [Miklós–Erdős–Soukup, Electron. J. Combin., 20, (2013)]

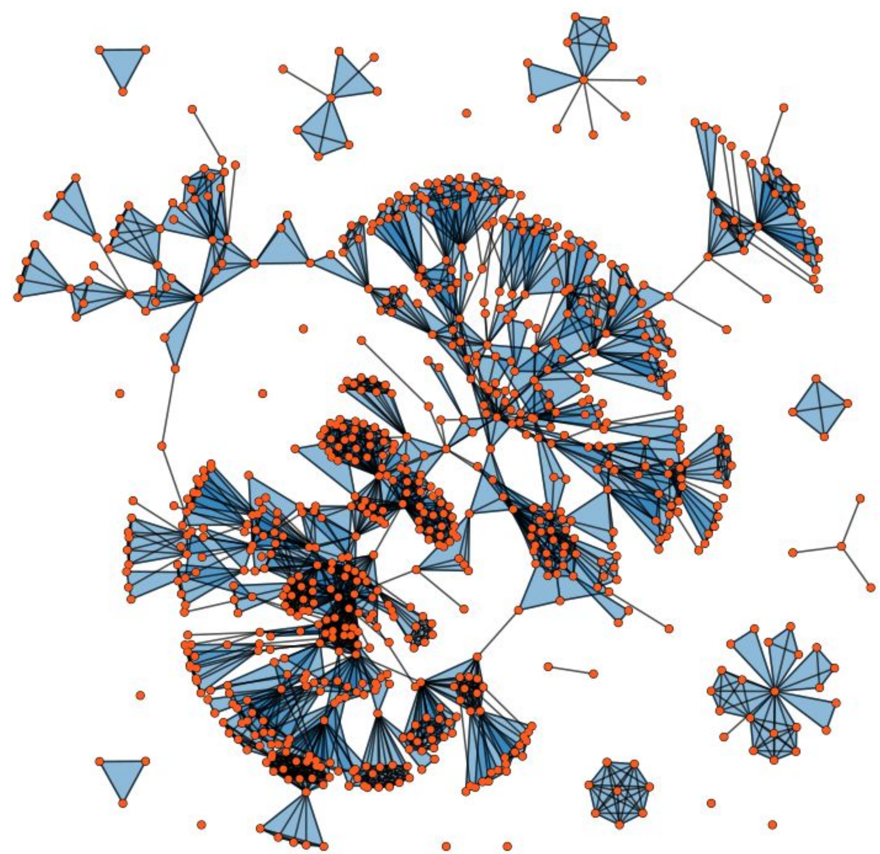
- MCMC is uniform over $\Omega(d, s)$
- Move set yields aperiodic chain
- Move set connects the space

Results - True systems

Disease regulation dataset
(facets : *genes*, nodes : *human diseases*)
[Goh et al., PNAS, **104**, (2007)]

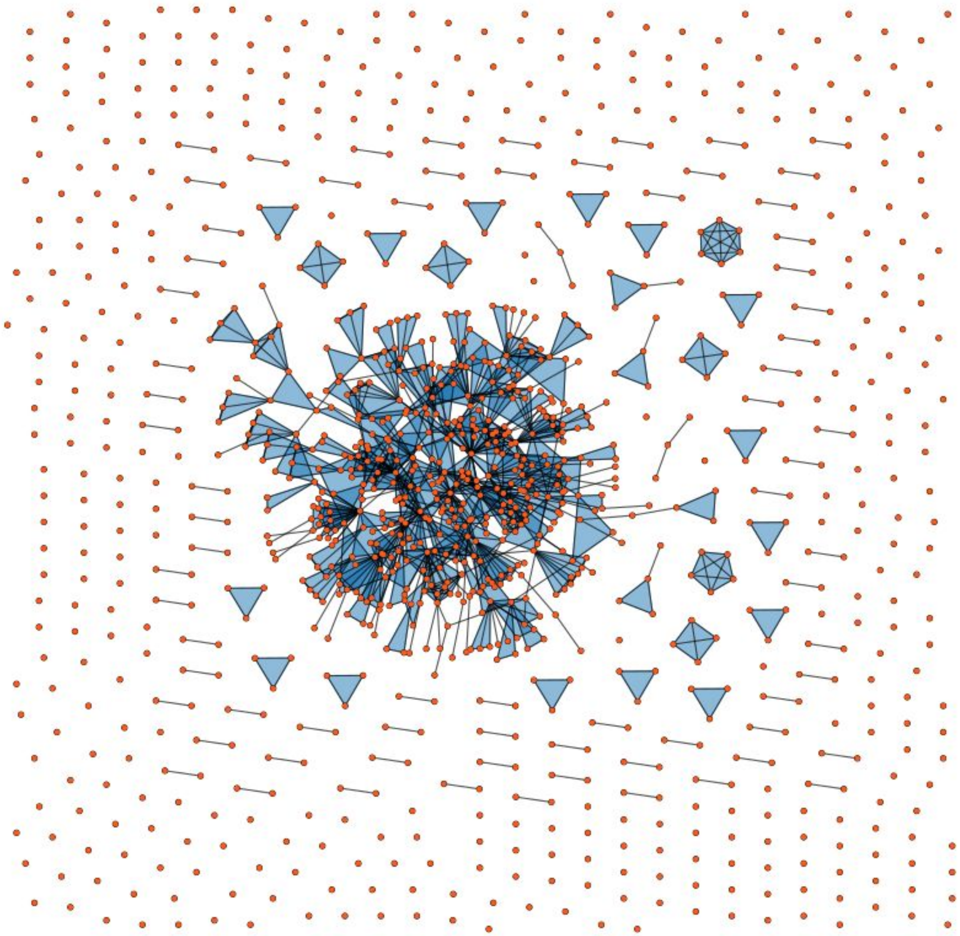


Crimes in St-Louis (true system)
(facets : *people*, nodes : *crimes*)
[Rosenfeld et al., (1991)]

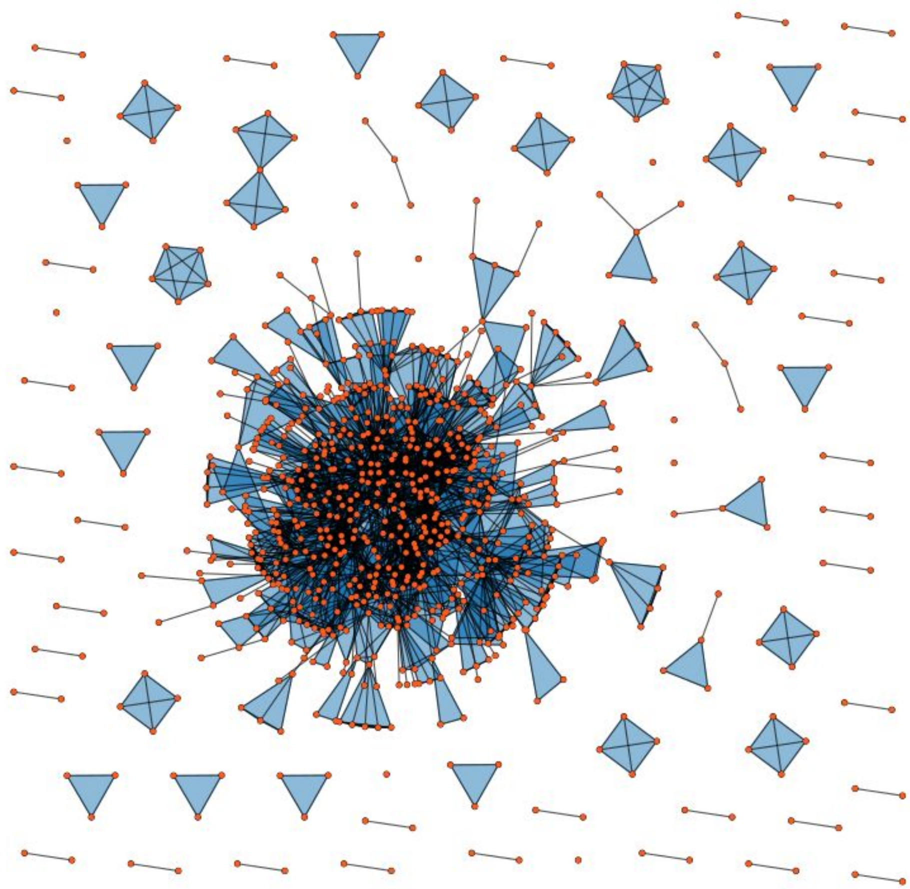


Results - Random instances

Disease regulation dataset (random instance)
(facets : *genes*, nodes : *human diseases*)
[Goh et al., PNAS, **104**, (2007)]



Crimes in St-Louis (random instance)
(facets : *people*, nodes : *crimes*)
[Rosenfeld et al., (1991)]



Indiana University
Network Science Institute

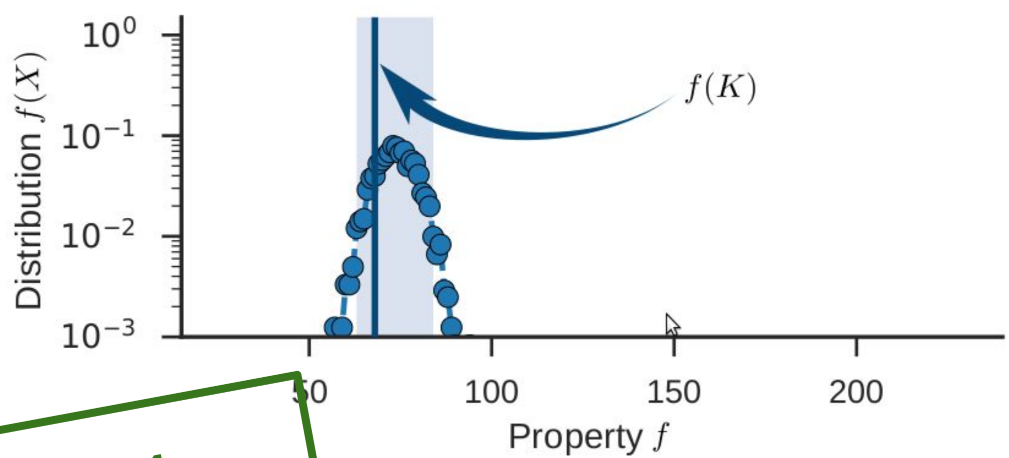
How to assess the significance of the properties of a simplicial complex?

Building a null model

Concept for a null model

Null model

Is the quantity **f(X)** close to **f(K)** for random simplicial complexes **X**~**SCM[d(K), s(K)]** ?



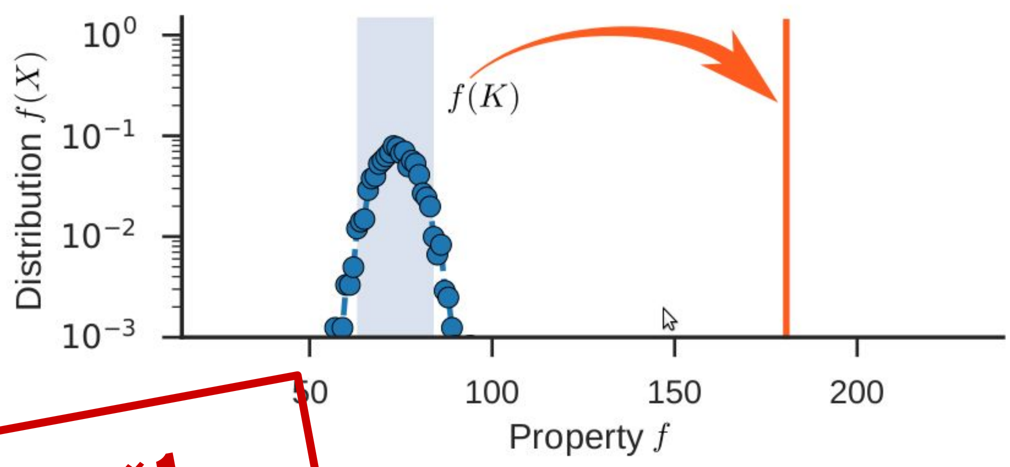
Pr[| f(K) − f(X)| <] ≈ 1

K is typical, the local quantities (***d,s***) explain **f**.

Concept for a null model

Null model

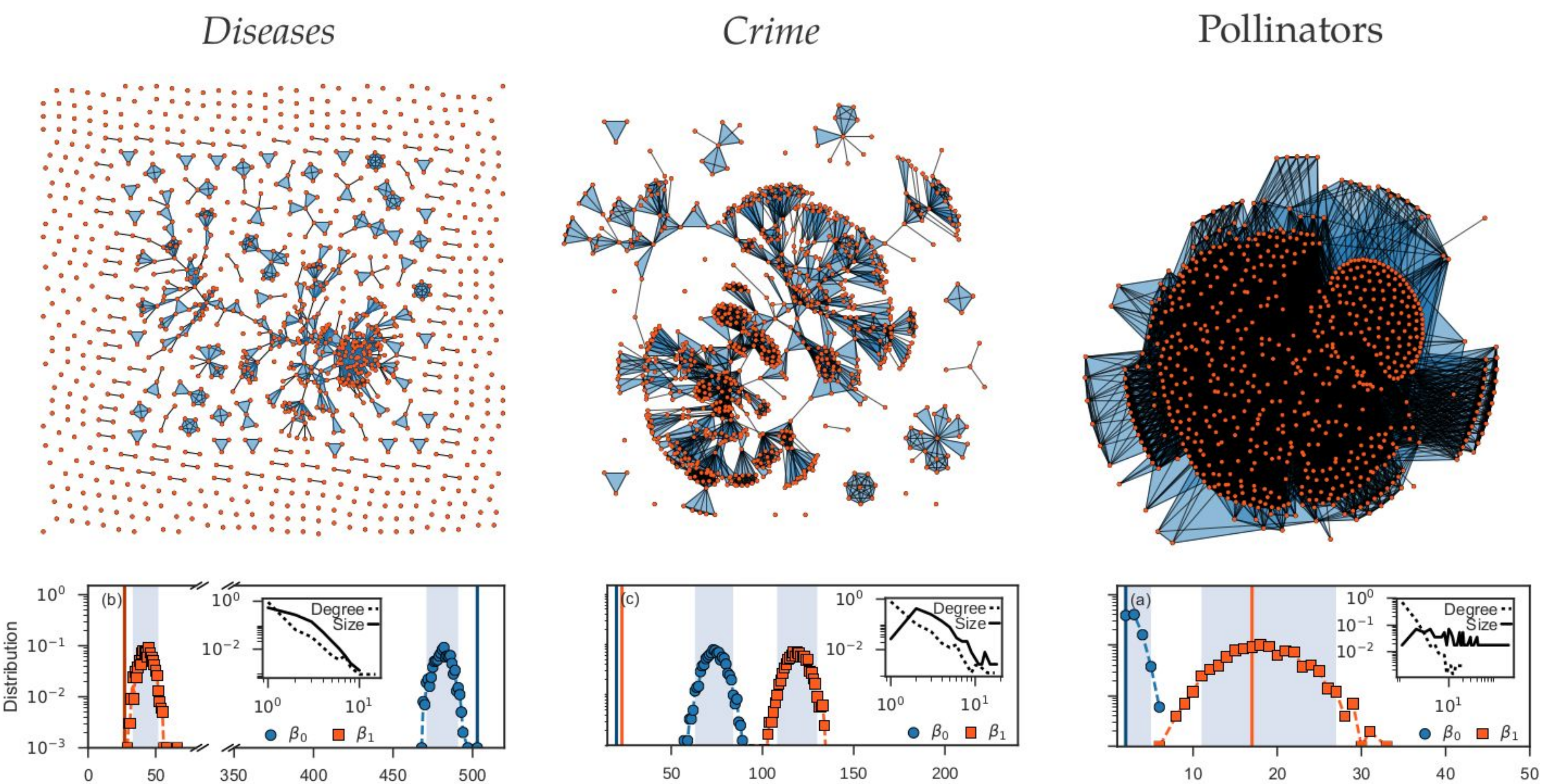
Is the quantity **f(X)** close to **f(K)** for random simplicial complexes **X~SCM[d(K), s(K)]** ?



Pr[| f(K) – f(X)| <] ≪ 1

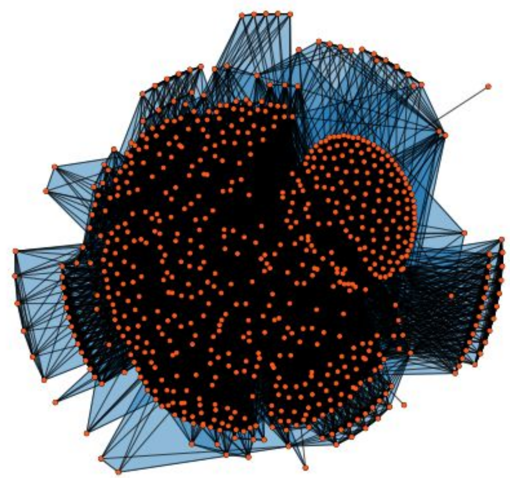
K is atypical, K is organized beyond the local scale.

Results on Betti numbers of real data sets

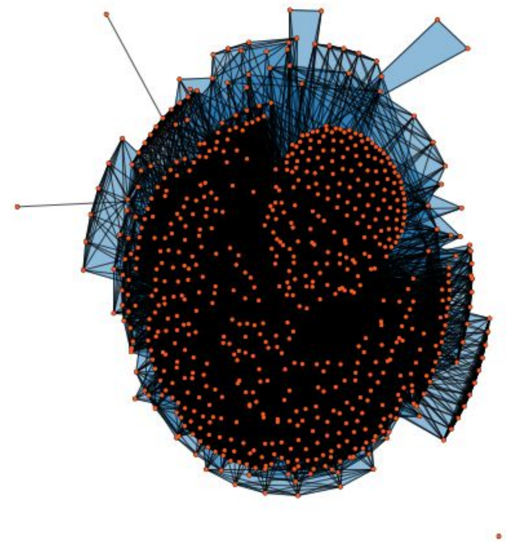


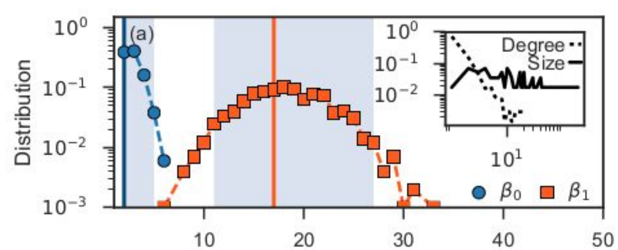
Results on Betti numbers of real data sets

Pollinators (*real*)



Pollinators (*random*)





Code : github.com/jg-you/scm

Our contribution

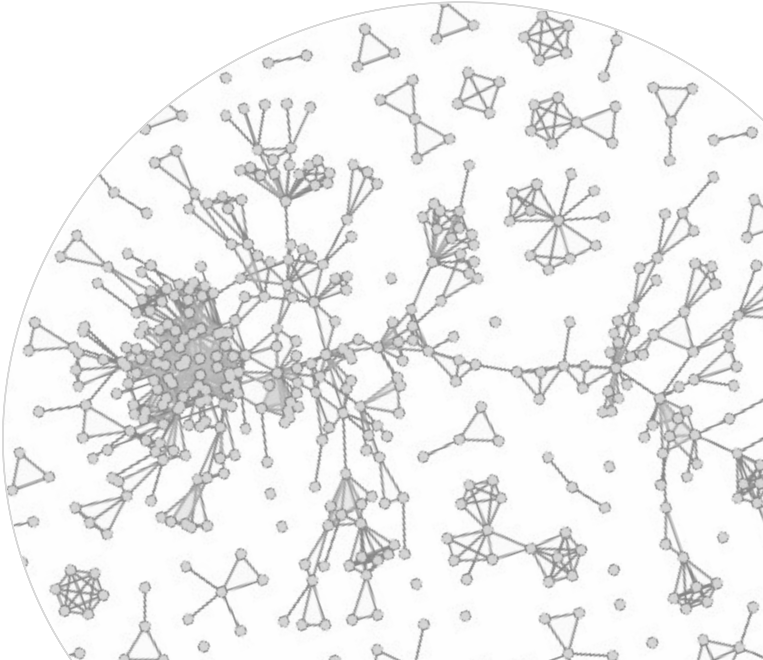
(1) J.-G. Young, G. Petri, F. Vaccarino and A. Patania, Phys. Rev. E 96 (3), 032312 (2017)

Equilibrium random ensembles

(2) O. Courtney and G. Bianconi, Phys. Rev. E 93, (2016)
(3) K. Zuev, O. Eisenberg and K. Krioukov, J. Phys. A 48, (2015)

Sampling

(4) B. K. Fosdick, et al., arXiv :1608.00607 (2016)

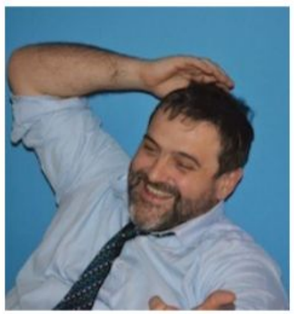


Indiana University
Network Science Institute

Our contribution

Code : github.com/jg-you/scm

(1) J.-G. Young, G. Petri, F. Vaccarino and A. Patania, Phys. Rev. E 96 (3), 032312 (2017)



Francesco Vaccarino

Giovanni Petri

Jean-Gabriel Young

Thank You for the Attention

INSTITUTE FOR SCIENTIFIC INTERCHANGE FOUNDATION

Politecnico di Torino 1859 · 1906

Compagnia di San Paolo

Fonds de recherche sur la nature et les technologies Québec

UNIVERSITÉ LAVAL