A model of homophilic growth in higher-order networks

Violet Ross
University of Colorado Boulder

A **higher-order network** is comprised of entities and the multi-way interactions between them.

A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

A **higher-order network** is comprised of entities and the multi-way interactions between them.

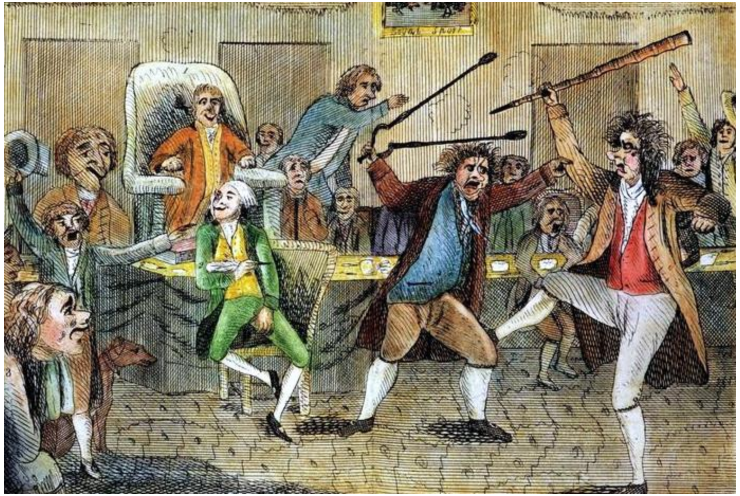
We can think of the United States Congress as a HON.

AP Photo/Andrew Harnik

A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

Collection of the U.S. House of Representatives



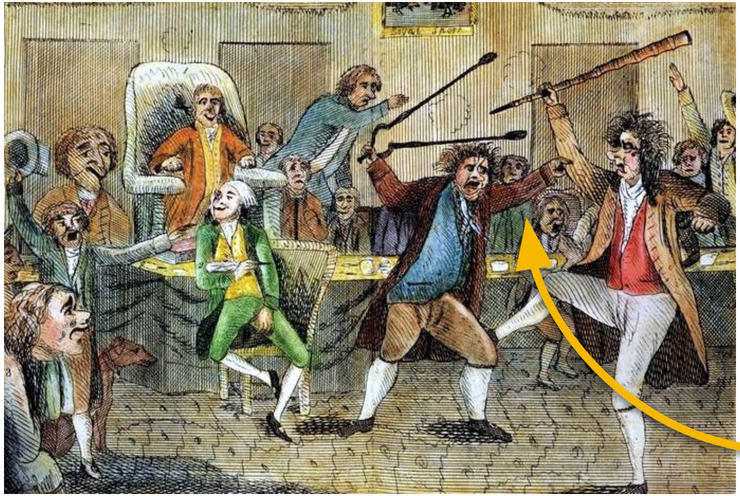
AP Photo/Andrew Harnik

A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

AP Photo/Andrew Harnik

Collection of the U.S. House of Representatives

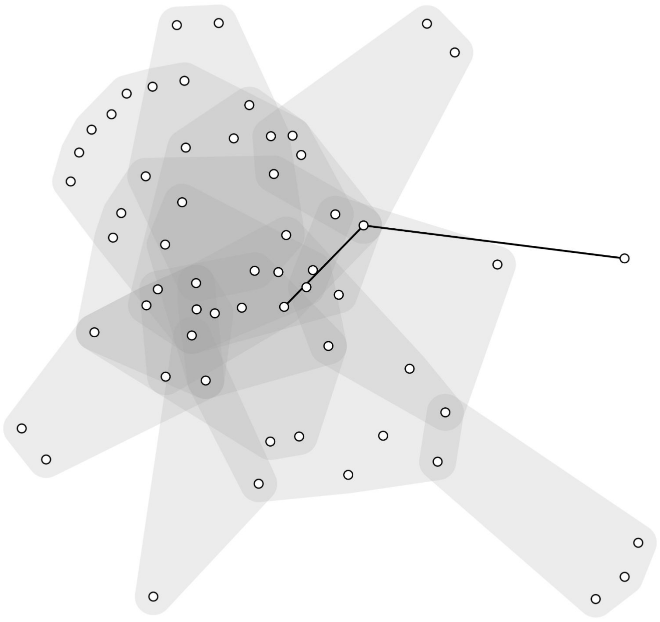


higher-order interaction!

A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

Hypergraphs are useful data structures for representing HONs.

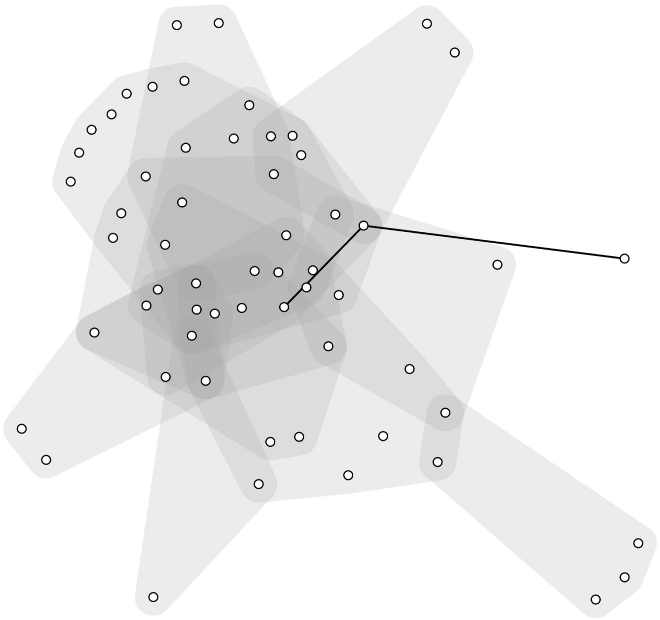


A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

Hypergraphs are useful data structures for representing HONs.

Some hypergraphs evolve over time.

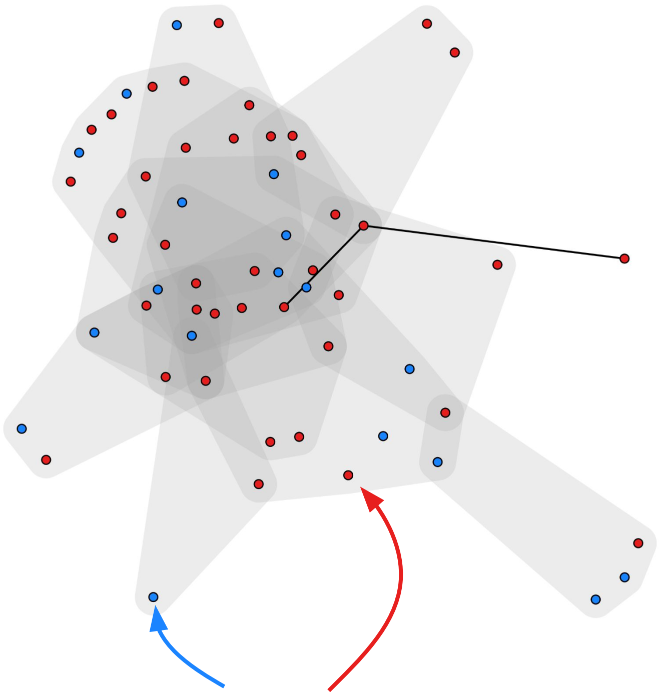


A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

Hypergraphs are useful data structures for representing HONs.

Some hypergraphs evolve over time.



And some of these time-evolving hypergraphs have binary node labels.

A **higher-order network** is comprised of entities and the multi-way interactions between them.

We can think of the United States Congress as a HON.

Hypergraphs are useful data structures for representing HONs.

Some hypergraphs evolve over time.

We wonder how homophily with respect to the node labels influences the evolution process.

homophily: like attracts like

We propose a model of homophilic hypergraph growth

We propose a model of homophilic hypergraph growth

This allows us to…

- **Investigate the dynamics** of homophilic growth in higher order networks.
- **Infer parameters** that describe homophily for real-world networks.

We propose a model of homophilic hypergraph growth

This allows us to…

- **Investigate the dynamics** of homophilic growth in higher order networks.
- **Infer parameters** that describe homophily for real-world networks.

big picture: understanding homophily as an aspect of HON evolution, rather than as a static property of a hypergraph in its final state

Team

Phil Chodrow
Middlebury College

Frannie Cataldo
Middlebury College

Team

Edge correlations and link prediction in growing hypergraphs

Xie He *
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755, USA
Microsoft. One Microsoft Way, Redmond, Washington 98052, USA

Philip S. Chodrow *,†
Department of Computer Science, Middlebury College, Middlebury, Vermont 05753, USA

Peter J. Mucha
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755, USA

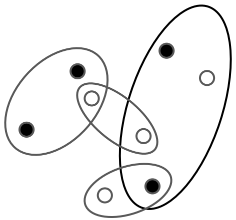
Phil Chodrow
Middlebury College

Frannie Cataldo
Middlebury College

The Model

At each timestep t:

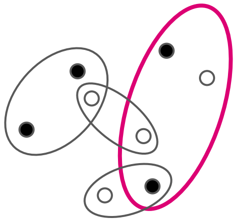
a)



At each timestep t:

a) Select an edge ***e*** uniformly at random.

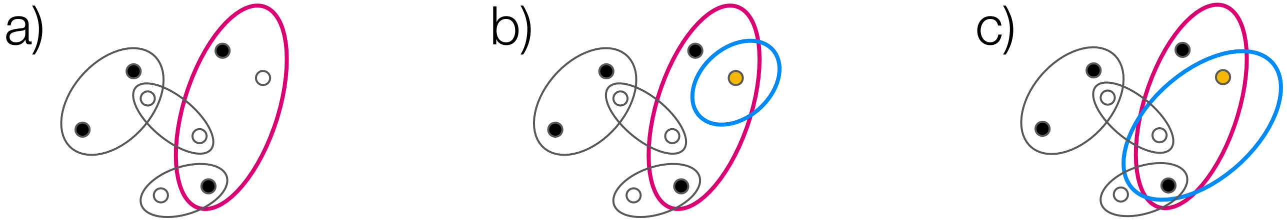
a)



b) Select a “pivot node” *u* in ***e*** uniformly at random. Create a new edge ***e’*** and add *u* to it.

The edge copy step

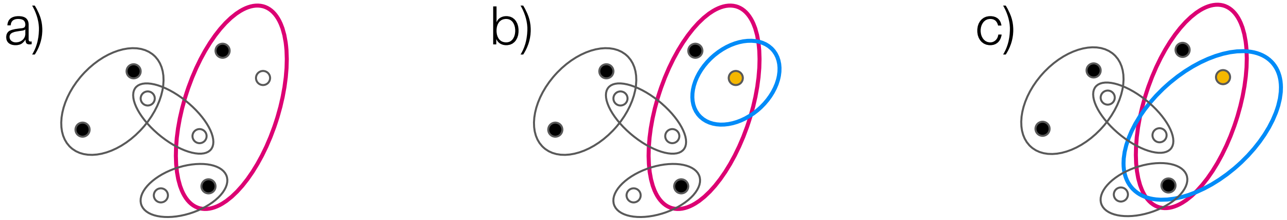
c) Add the other nodes from *e* to *e’* with probability ***p*** if they share the same node label as *u* and probability ***q*** otherwise.



The edge copy step

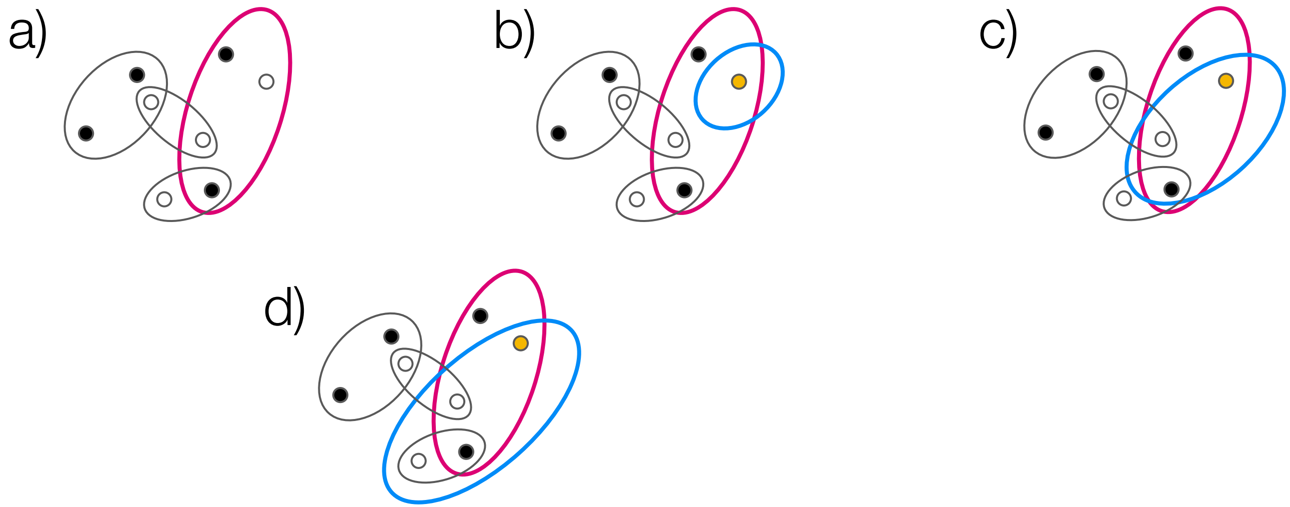
c) Add the other nodes from *e* to *e’* with probability $\boldsymbol{p}$ if they share the same node label as *u* and probability $\boldsymbol{q}$ otherwise.

Homophily arises when $p > q$



The external node addition step

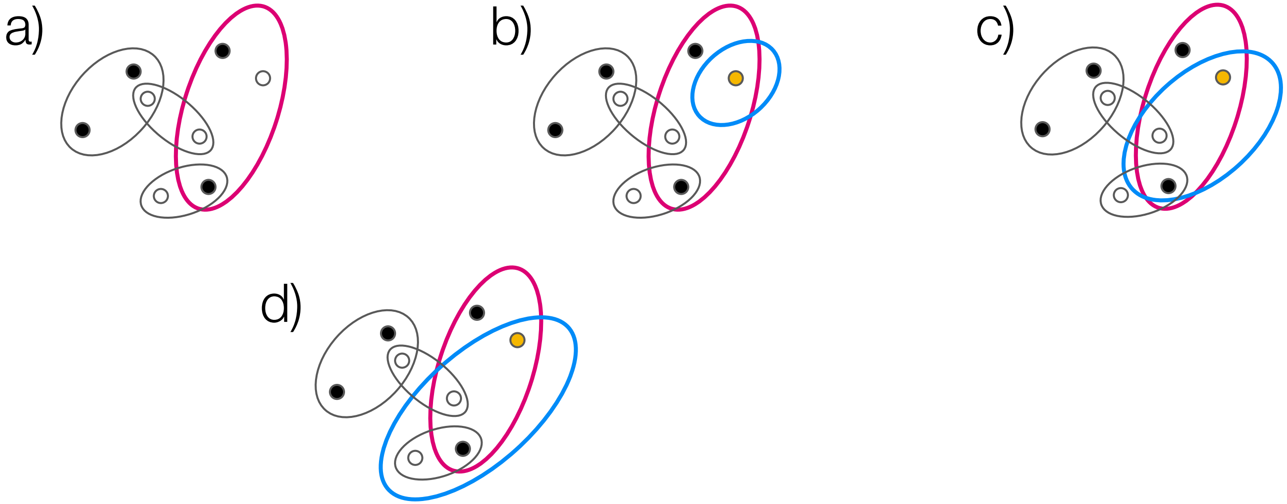
d) Generate a quantity from a Poisson distribution with parameter γ_{e,z_u} and add that many nodes outside of *e* with the same label as *u* to *e'*. Generate a value from a Poisson with parameter $\gamma_{e,\bar{z}_u}$ and add that many outside of *e* with the opposite label.



The external node addition step

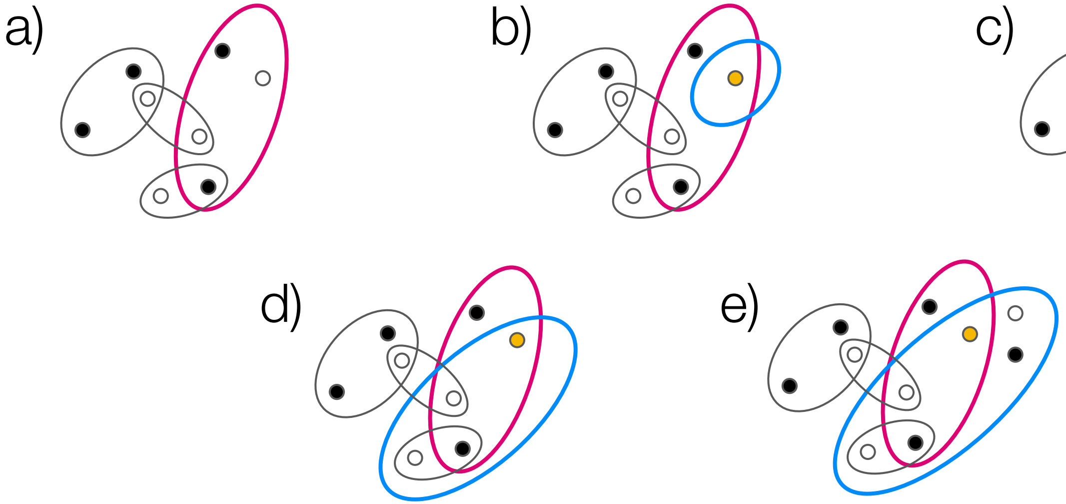
d) Generate a quantity from a Poisson distribution with parameter γ_{e,z_u} and add that many nodes outside of **e** with the same label as *u* to **e'**. Generate a value from a Poisson with parameter $\gamma_{e,\bar{z}_u}$ and add that many outside of **e** with the opposite label.

Homophily arises when $\gamma_{e,z_u} > \gamma_{e,\bar{z}_u}$



The novel node addition step

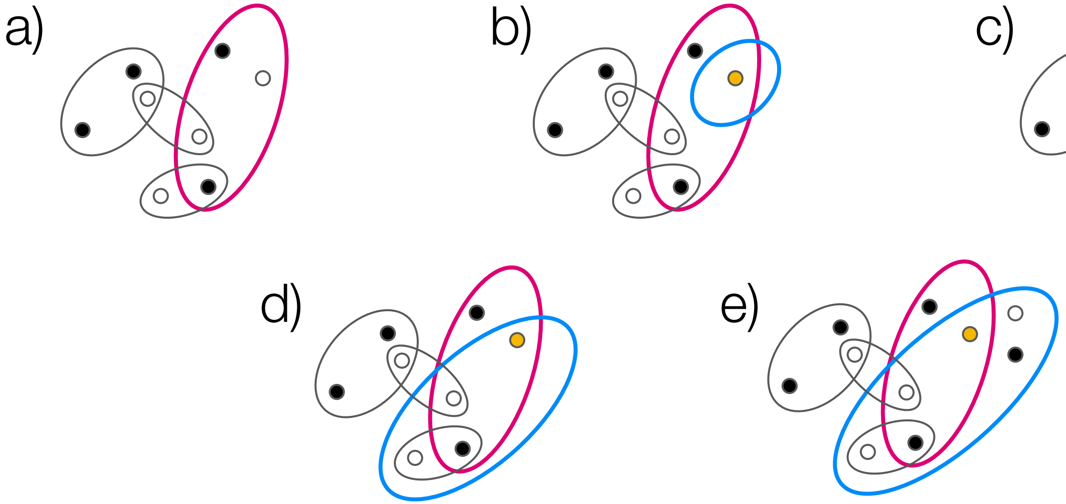
e) Generate a quantity from a Poisson distribution with parameter γ_{n,z_u} and add that many new nodes to the hypergraph, assign them the same label as *u*, and add them to to ***e'***. Do the same process to add new nodes of the opposite label, using a Poisson with parameter $\gamma_{n,\bar{z}_u}$.



The novel node addition step

e) Generate a quantity from a Poisson distribution with parameter γ_{n,z_u} and add that many new nodes to the hypergraph, assign them the same label as *u*, and add them to to *e'*. Do the same process to add new nodes of the opposite label, using a Poisson with parameter $\gamma_{n,\bar{z}_u}$.

Homophily arises when $\gamma_{n,z_u} > \gamma_{n,\bar{z}_u}$



Dynamics

Dynamics

AKA: what can we say about a hypergraph grown using this model?

Power Law Degree Distribution

Assume that the degree of each node is independent of the ratio of 0-labeled nodes in the edge that was sampled to form it.

The degree distribution of a hypergraph generated by our model follows a power law.

We can derive its exponent in terms of our model parameters.

$$\zeta = 1 + \frac{\langle k \rangle}{1 + p(\rho_{00} + \rho_{11} - 1) + 2q\rho_{01}}$$

$$\rho_{ij} = \left\langle \frac{k_i k_j}{k} \right\rangle$$

$\langle \ \rangle$: mean

k : number of nodes in edge

k_i : number of nodes of label *i* in edge

Power Law Degree Distribution

Assume that the degree of each node is independent of the ratio of 0-labeled nodes in the edge that was sampled to form it.

The degree distribution of a hypergraph generated by our model follows a power law.

We can derive its exponent in terms of our model parameters.

$$\zeta = 1 + \frac{\langle k \rangle}{1 + p(\rho_{00} + \rho_{11} - 1) + 2q\rho_{01}}$$

$$\rho_{ij} = \left\langle \frac{k_i k_j}{k} \right\rangle$$

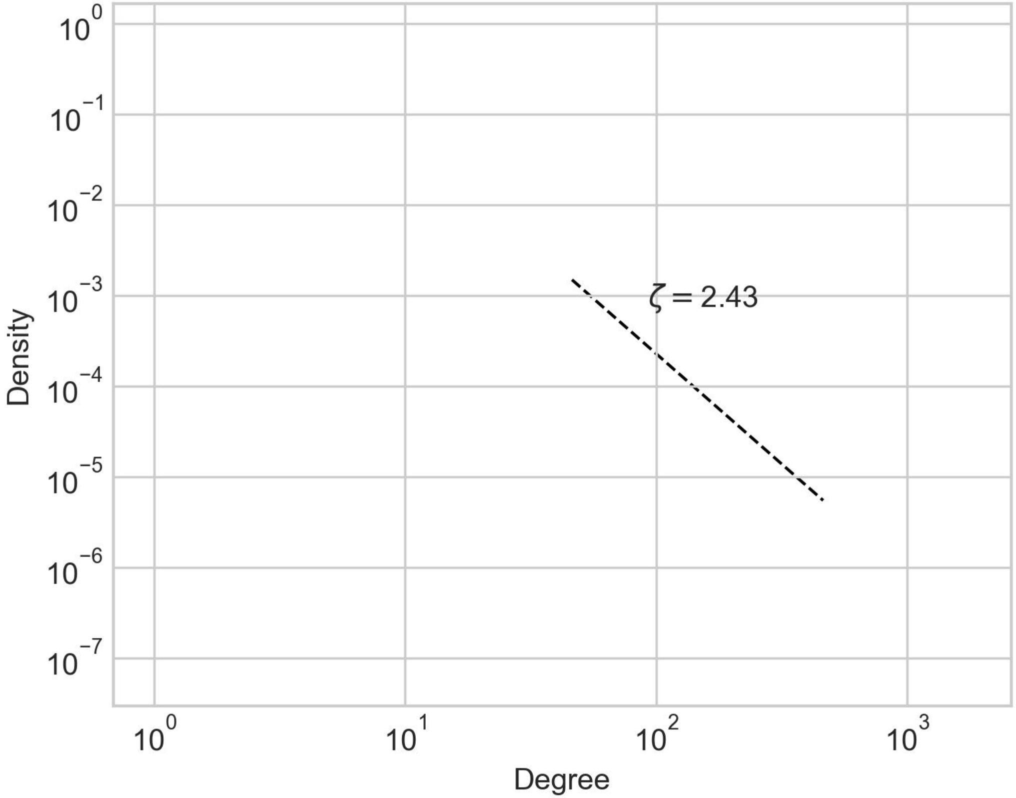
p, q: copy parameters

$\langle \ \rangle$: mean

k : number of nodes in edge

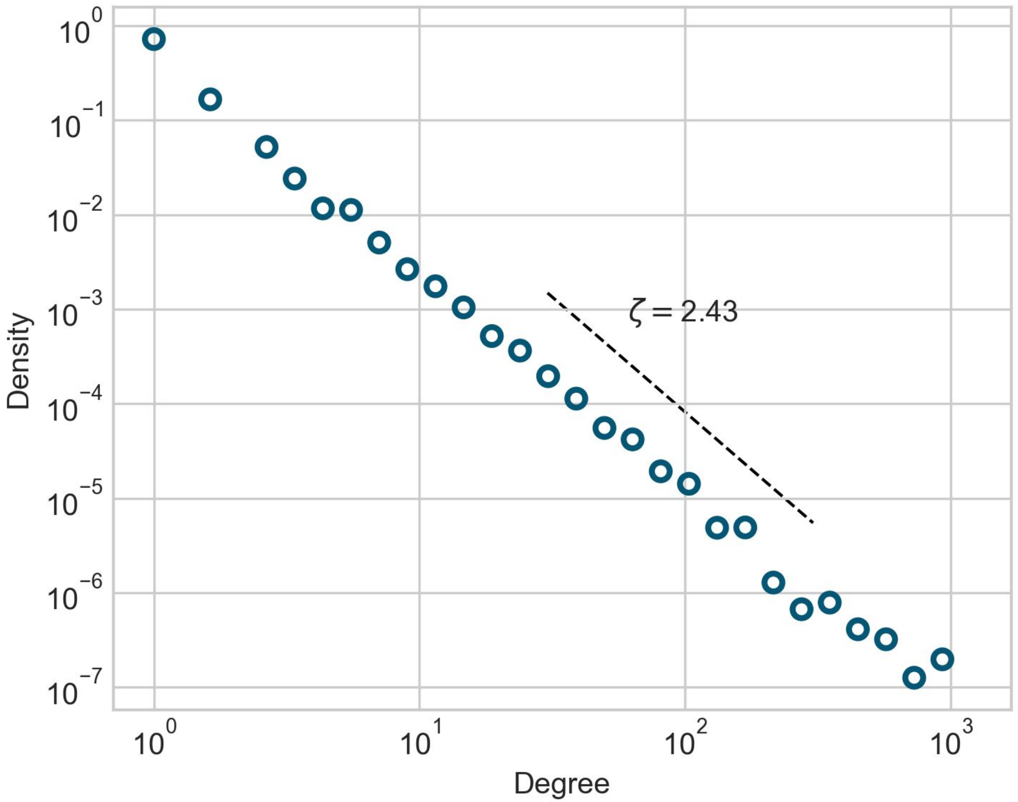
k_i : number of nodes of label *i* in edge

Power Law Degree Distribution



$p = 0.7,\ q = 0.3,$
$\gamma_{e,z_u} = 0.5,\ \gamma_{e,\bar{z}_u} = 0.3,$
$\gamma_{n,z_u} = 0.2,\ \gamma_{n,\bar{z}_u} = 0.1$

Power Law Degree Distribution



$p = 0.7$, $q = 0.3$,
$\gamma_{e,z_u} = 0.5$, $\gamma_{e,\bar{z}_u} = 0.3$,
$\gamma_{n,z_u} = 0.2$, $\gamma_{n,\bar{z}_u} = 0.1$

Inference

Given a hypergraph, can we determine what parameters generated it?

How **homophilic** is the formation of new connections?

How strong is the edge copy **mechanism**?

Maximum Likelihood Estimation

Maximizes over the log marginal likelihood by **direct optimization of the likelihood function**

Maximum Likelihood Estimation

Maximizes over the log marginal likelihood by **direct optimization of the likelihood function**

Maximum Likelihood Estimation

Maximizes over the log marginal likelihood by **direct optimization of the likelihood function**

Expectation Maximization

Maximizes over the log marginal likelihood by **computing expected sufficient statistics**

Avoids marginalizing over latent variables

Maximum Likelihood Estimation

Maximizes over the log marginal likelihood by **direct optimization of the likelihood function**

Expectation Maximization

Maximizes over the log marginal likelihood by **computing expected sufficient statistics**

Avoids marginalizing over latent variables

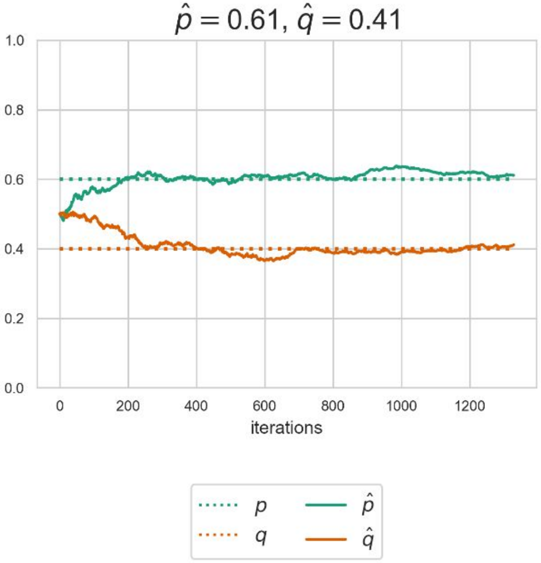
Stochastic Expectation Maximization

Maximizes over the log marginal likelihood by **computing expected sufficient statistics** ***given a single observation***

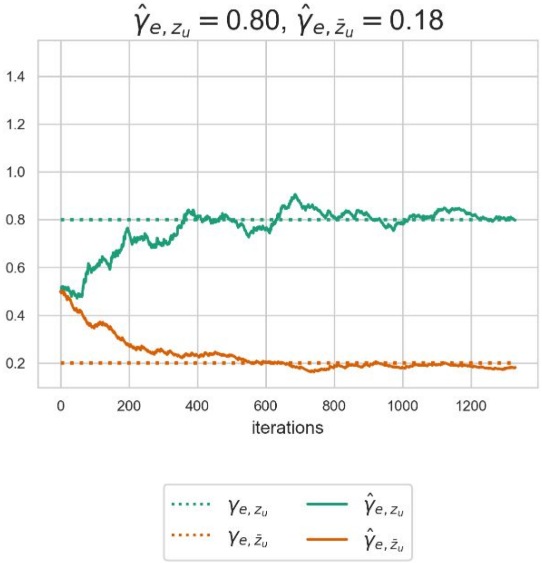
Avoids marginalizing over latent variables and computing expectations over the complete data

SEM performs well on synthetic data

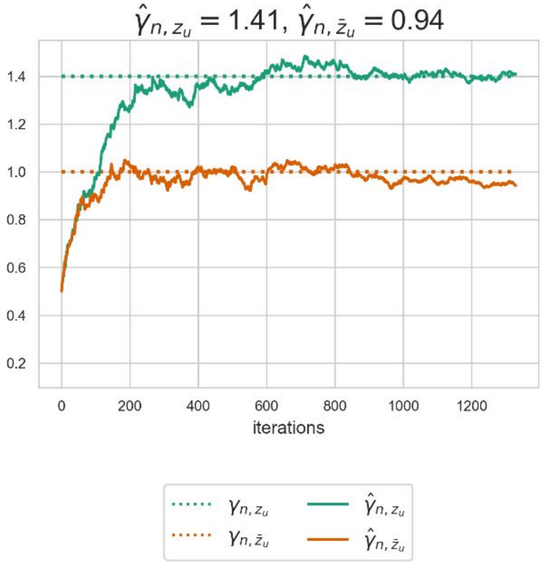
Edge copy step



External node step



Novel node step

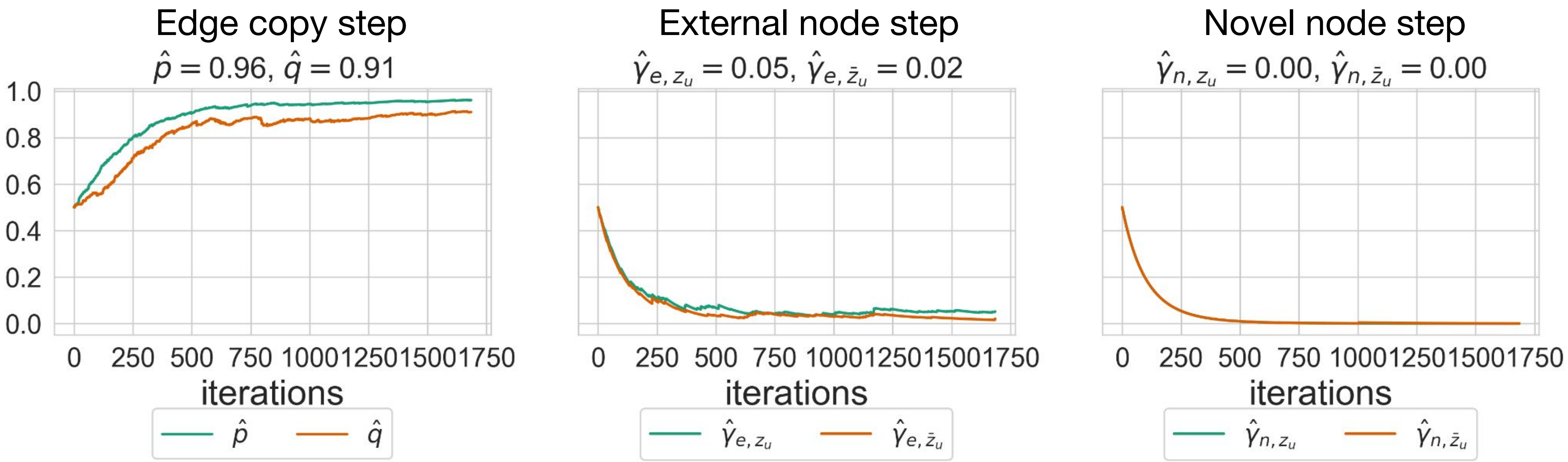


Inference using our model tells us about the properties of real-world higher-order networks.

Some hypergraphs exhibit little to no homophily when forming new connections

data: social interactions between high school students

binary labeling: gender M/F



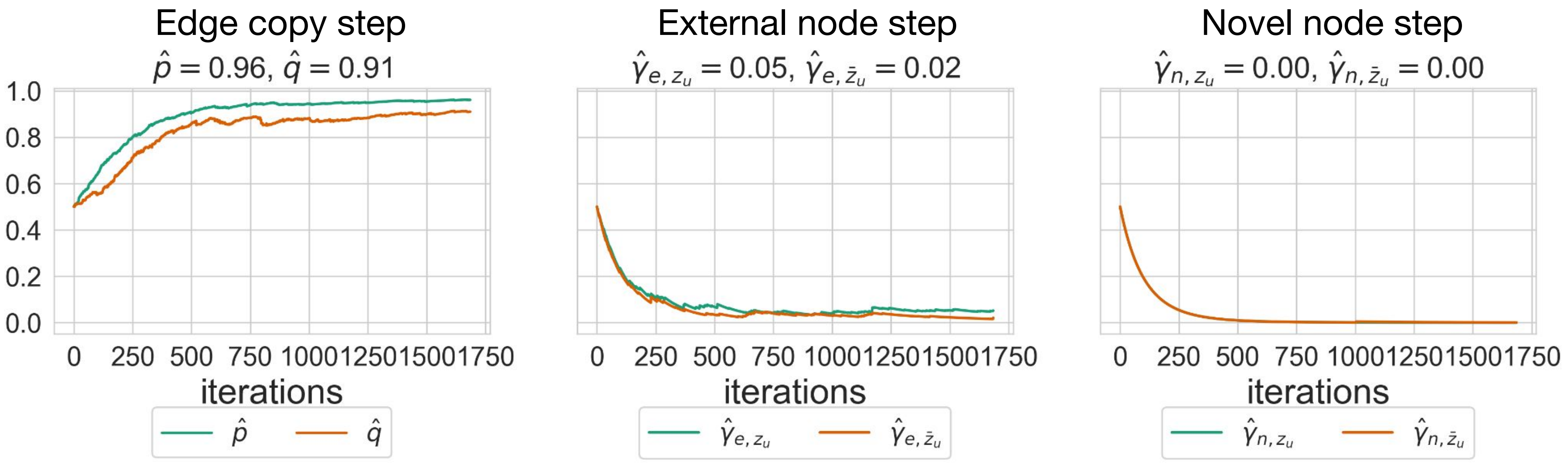
Fournet, Julie, and Alain Barrat. "Contact patterns among high school students." *PloS one* 9.9 (2014): e107878.

Some hypergraphs exhibit little to no homophily when forming new connections

data: social interactions between high school students

binary labeling: gender M/F

Note #1: This result is dependent on the choice of label



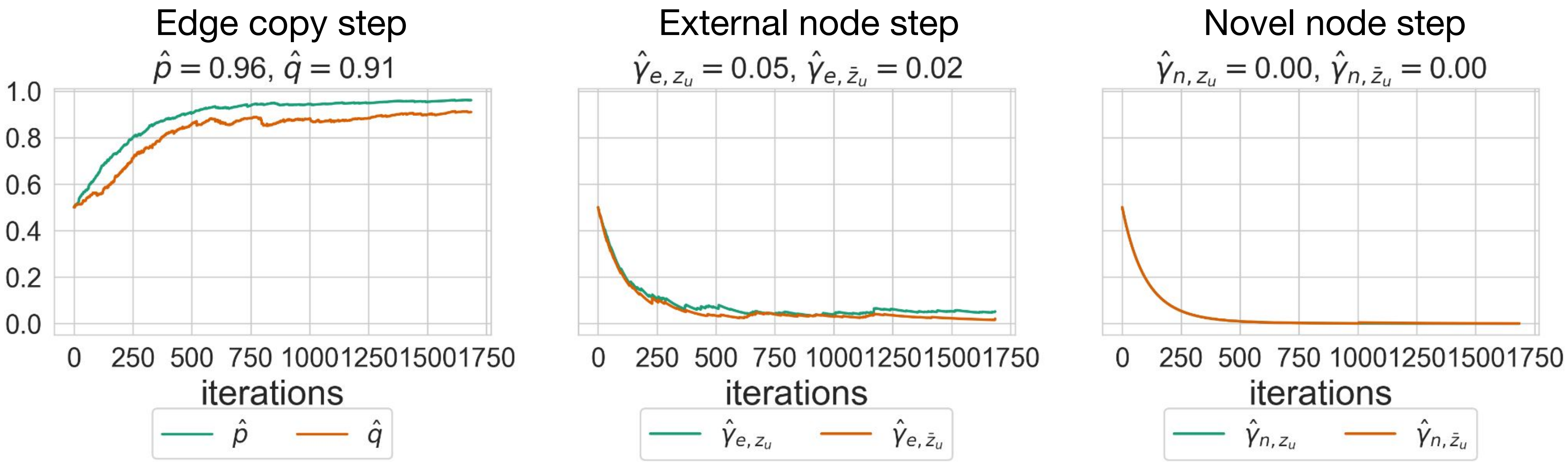
Fournet, Julie, and Alain Barrat. "Contact patterns among high school students." *PloS one* 9.9 (2014): e107878.

Some hypergraphs exhibit little to no homophily when forming new connections

data: social interactions between high school students

binary labeling: gender M/F

Note #2: Lack of homophily ≠ equal representation (mean proportion of the within-edge majority is 0.85)

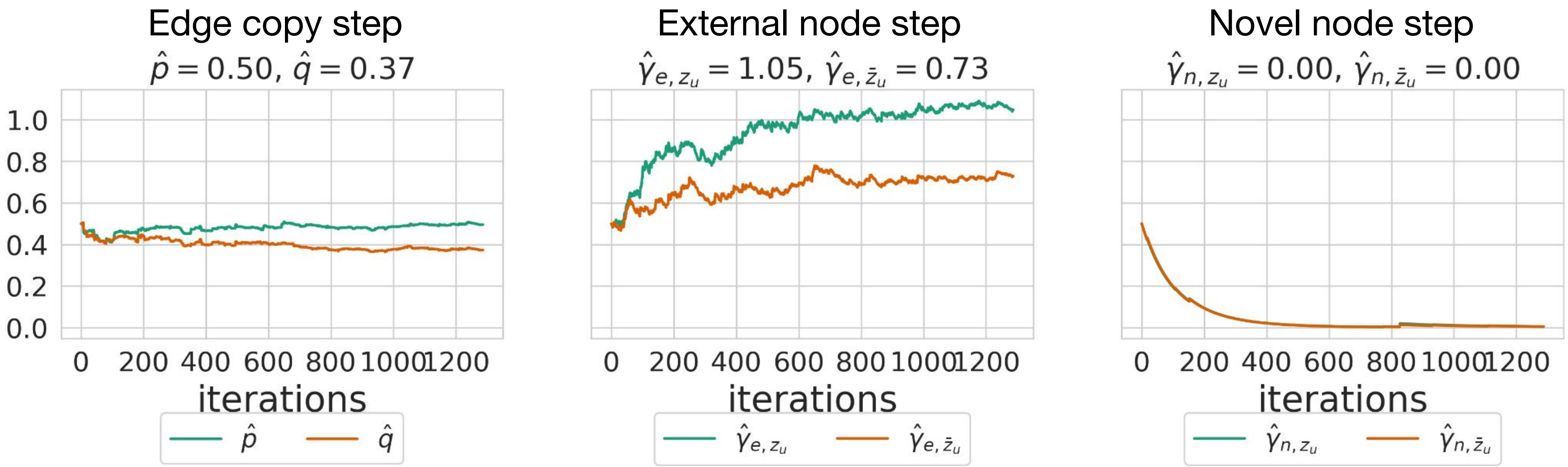


Fournet, Julie, and Alain Barrat. "Contact patterns among high school students." *PloS one* 9.9 (2014): e107878.

Homophily is evident in the growth processes of others

data: bill cosponsorship in the United States Senate

binary labeling: political party D/R



Fowler, James H. "Legislative cosponsorship networks in the US House and Senate." *Social networks* 28.4 (2006): 454-465., Fowler, James H. "Connecting the congress: A study of cosponsorship networks." *Political analysis* 14.4 (2006): 456-487., Chodrow, Philip S., Nate Veldt, and Austin R. Benson. "Generative hypergraph clustering: From blockmodels to modularity." *Science Advances* 7.28 (2021).

Conclusions

We proposed a model of hypergraph growth where homophily can appear in three different ways:

1. Edge copying
2. External node addition
3. Novel node addition

We can describe the dynamics of hypergraphs that grow in this way:

- They have power law distributed node degrees with known exponent

We can infer the model parameters using stochastic expectation maximization.

- This provides insights into the evolution dynamics of real-world higher order networks.

Discussion

There are many ways to investigate the properties of a higher-order network with respect to its node labels.

Often, these tell us about the **structure** of a labeled hypergraph in its **final state**

Our model informs on the **mechanisms** that produce that **final hypergraph**

So, in addition to saying “this network has community structure”

we can also say “this is how the community structure emerged”

Thanks + Acknowledgements

Phil Chodrow
Middlebury College

Frannie Cataldo
Middlebury College

Dan Larremore
University of Colorado Boulder

My website