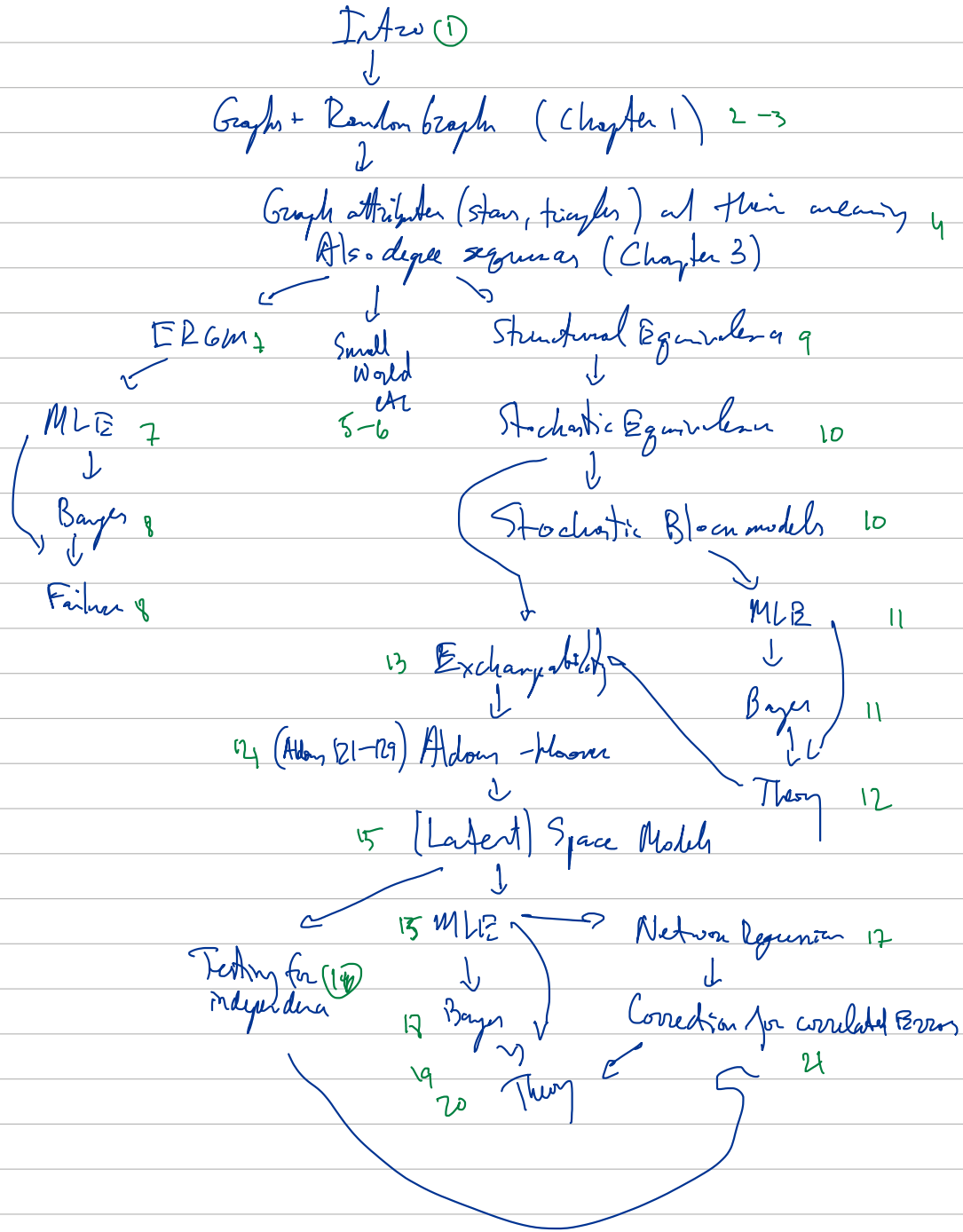
Theory and Methods for the Analysis of Social Networks

Alexander Volfovsky
Department of Statistical Science, Duke University

Lecture 1: January 16, 2018


Intro (1)
Graphs + Random graphs (Chapter 1) 2-3
Graph attributes (stars, triangles) and their meaning 4
Also degree sequences (Chapter 3)
ERGMs 7
Small World 5-6
Structural Equivalence 9
MLE 7
Stochastic Equivalence 10
Bayes 8
Stochastic Block models 10
Failure 8
MLE 11
13 Exchangeability
Bayes 11
14 (Aldous 121-129) Aldous-Hoover
Theory 12
15 [Latent] Space Models
13 MLE
Network Regression 17
Testing for independence (14)
17 Bayes
Correction for correlated Errors 21
19 20 Theory

Outline

Jan 11 : Brief intro and Guest lecture by James Moody, Duke Sociology
Jan 16 : Intro, why we do this
Jan 18 : Graph theory and random graphs
Jan 23 : Graph theory and random graphs
Jan 25 : Graph attributes
Jan 30 : Small world networks
Feb 1 : Small world networks
Feb 6 : Exponential Random Graph Models (Intro and MLE)
Feb 8 : Exponential Random Graph Models (Bayes and failures)
Feb 13 : Structural Equivalence
Feb 15 : Stochastic Equivalence and intro to stochastic blockmodels
Feb 20 : Stochastic blockmodels and the latent space model (MLE and Bayes)
Feb 22 : Stochastic blockmodels (theory)
Feb 27 : Stochastic blockmodels and belief propagation
March 1 : Aldous-Hoover theorem and the latent space model

Outline

March 6 : (catch up)
March 8 : Revisiting why we do this — applied examples
March 13 : [Spring Break]
March 15 : [Spring Break]
March 20 : Latent Space Models (MLE)
March 22 : Testing for independence
March 27 : Network regression
March 29 : Bayesian approaches to latent space models
April 3 : Bayesian approaches to latent space models
April 5 : Theory for latent space models
April 10 : Network regression with correlated errors

Class format

- Lectures — some fundamentals

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

- Duke Network Analysis Center seminars: `https://dnac.ssri.duke.edu`

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

- Duke Network Analysis Center seminars: `https://dnac.ssri.duke.edu`
- Assignments: several homeworks throughout the semester and a final project.

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

- Duke Network Analysis Center seminars: `https://dnac.ssri.duke.edu`
- Assignments: several homeworks throughout the semester and a final project.
- Course page

Class format

- Lectures — some fundamentals
- Case studies — interesting examples of network analysis as it is used

  e.g. Class 0: James Moody
- Lab sections — cover some additional material and all of the computing

- Duke Network Analysis Center seminars: `https://dnac.ssri.duke.edu`
- Assignments: several homeworks throughout the semester and a final project.
- Course page
- Online discussion: on Slack

Course goals

- Interested in understanding the formation of relationships

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?
  - How to develop fail-safes to overcome these problems?

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?
  - How to develop fail-safes to overcome these problems?
- Interested in implementation and methodology:

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?
  - How to develop fail-safes to overcome these problems?
- Interested in implementation and methodology:
  - How do we quickly estimate model parameters?

Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?
  - How to develop fail-safes to overcome these problems?
- Interested in implementation and methodology:
  - How do we quickly estimate model parameters?
  - How do we interpret model parameters when the model is wrong?

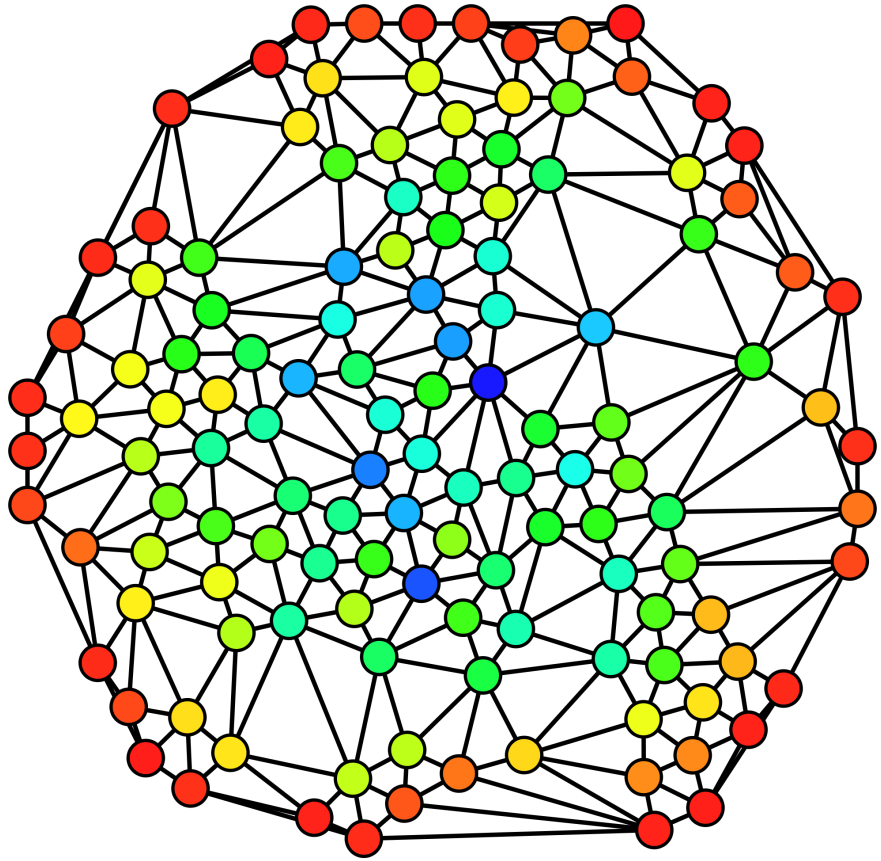
Course goals

- Interested in understanding the formation of relationships
- Applied fields: sociology, economics, biology, epidemiology
- Interested in fundamental theory questions:
  - What assumptions are made for different network models?
  - What models work when the assumptions fail?
  - How to develop fail-safes to overcome these problems?
- Interested in implementation and methodology:
  - How do we quickly estimate model parameters?
  - How do we interpret model parameters when the model is wrong?
  - How do we run experiments on networks?

Recent work

- Lots and lots of causal inference
- Big(gest) problem in causal inference: we assume that everything is independent.
- Reality: nothing is independent!

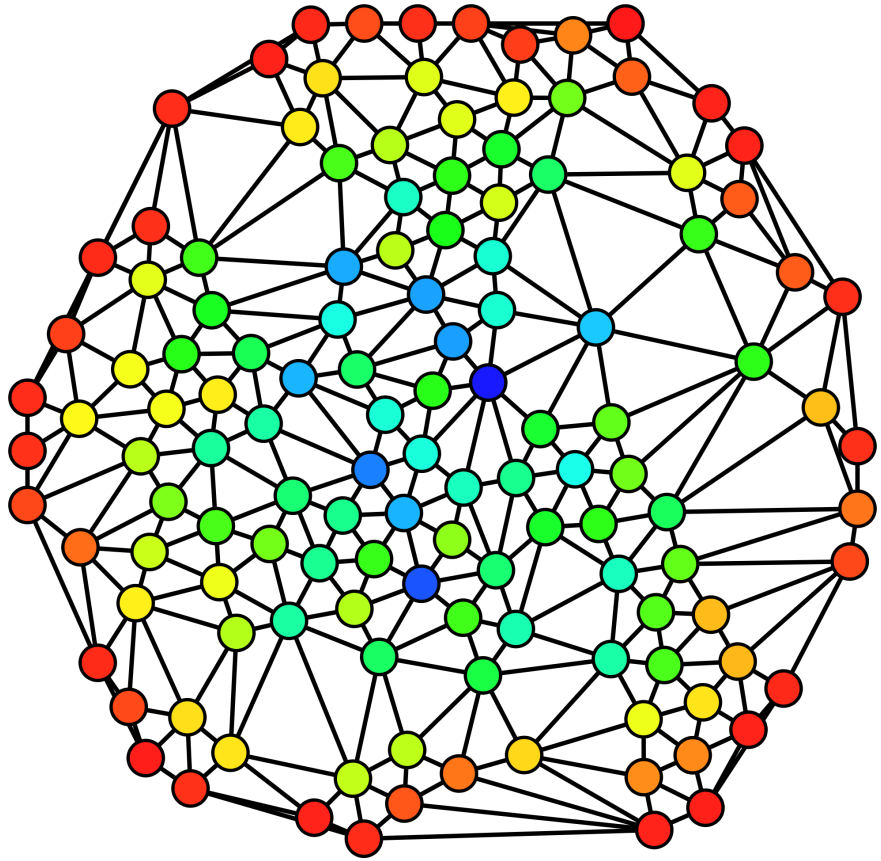
Some context: Facebook



- Facebook wants to change its' ad algorithm.

Source: Wikimedia

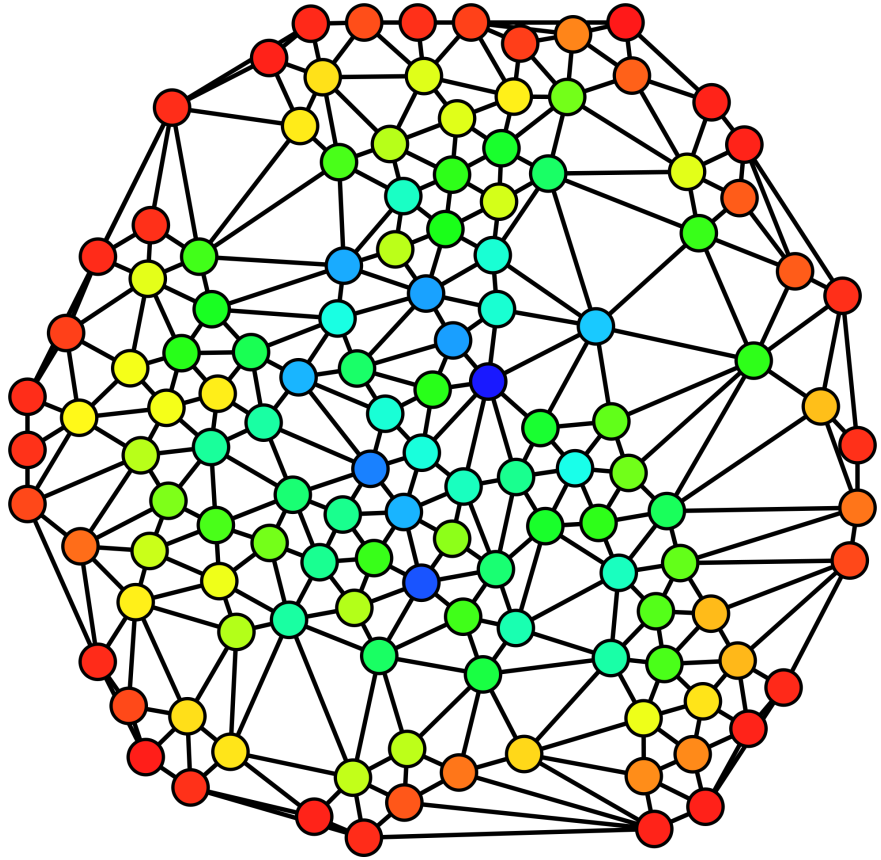
Some context: Facebook



- Facebook wants to change its' ad algorithm.
- Can't do it on the whole graph

Source: Wikimedia

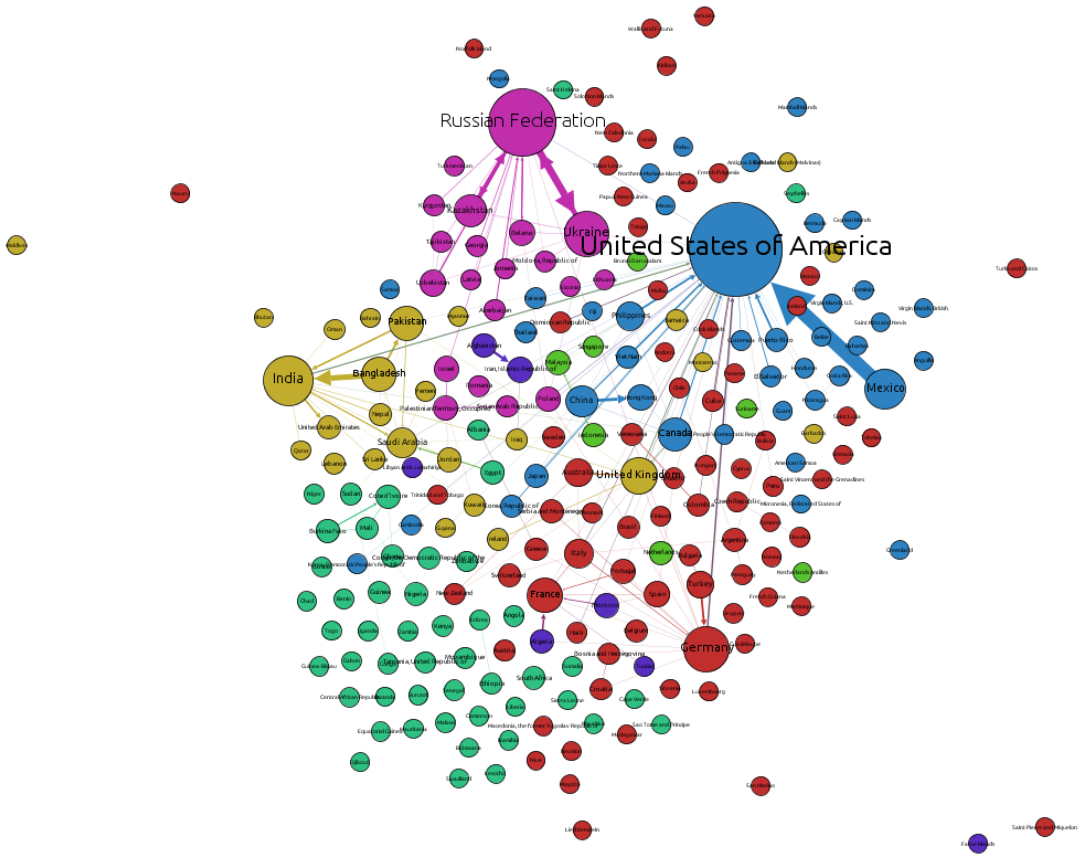
Some context: Facebook



- Facebook wants to change its' ad algorithm.
- Can't do it on the whole graph
- Need "total network effect"

Source: Wikimedia

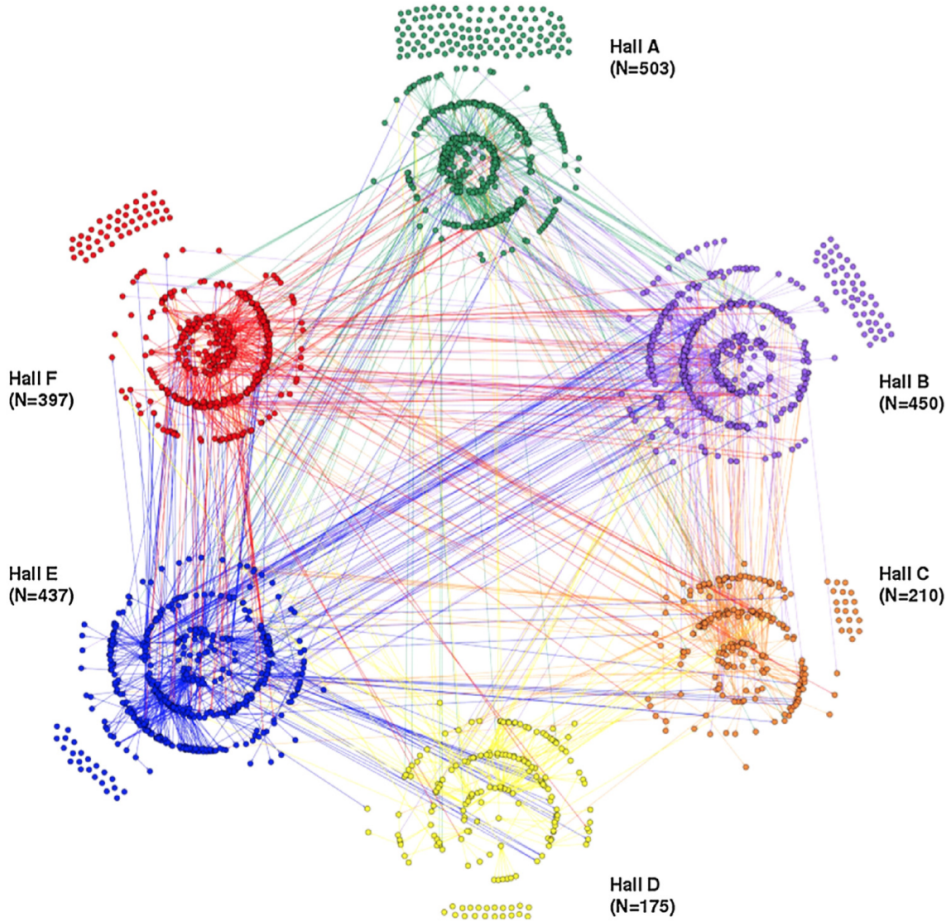
Some context: (im)migration



- Want to know how regime change affects population.
- Politicians during election years care about direct effects.

Source: http://openscience.alpine-geckos.at/courses/social-network-analyses/empirical-network-analysis/

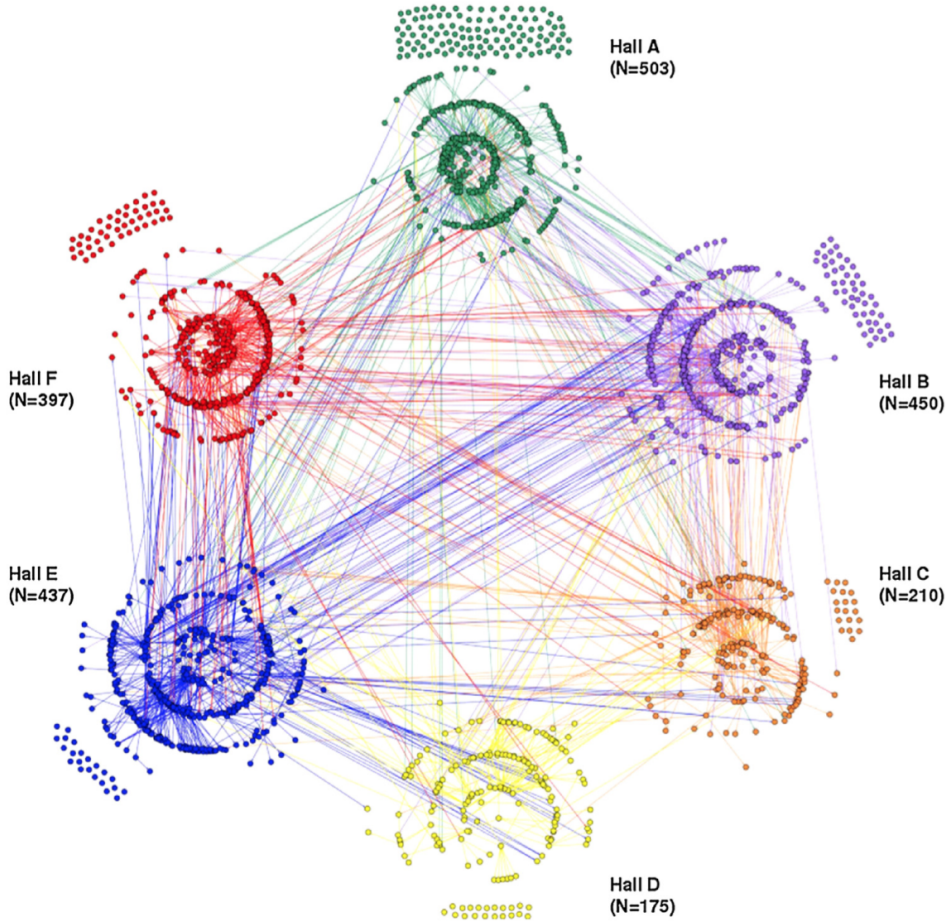
Some context: disease spread



- Want to study efficacy of isolation as treatment for influenza-like illness.

Source: Figure 9 of "Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial" by Aiello et al.

Some context: disease spread

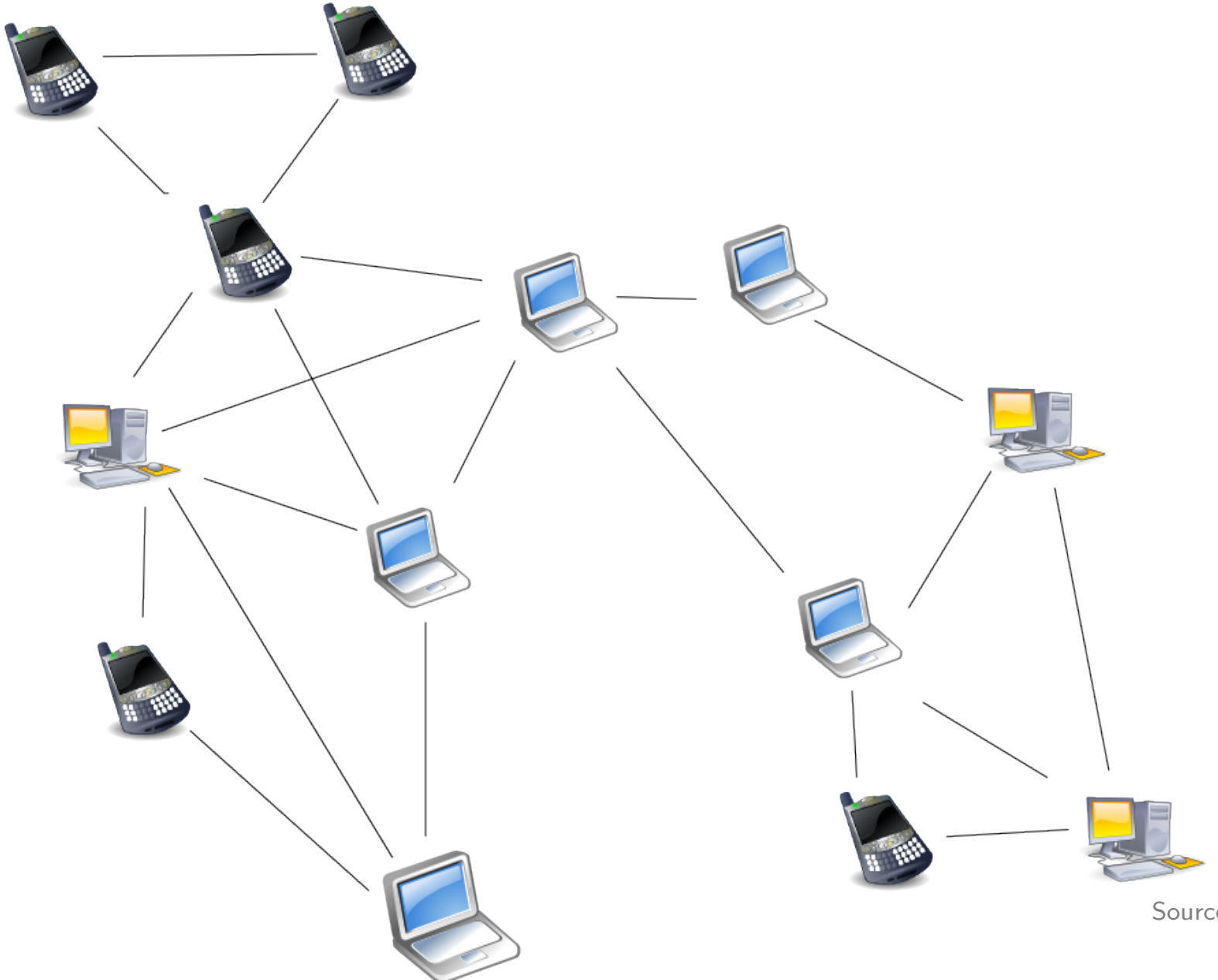


- Want to study efficacy of isolation as treatment for influenza-like illness.
- Interested in spread, duration of illness, etc.

Source: Figure 9 of "Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial" by Aiello et al.

Other network contexts

Studying computer network congestion



Source: Wikimedia

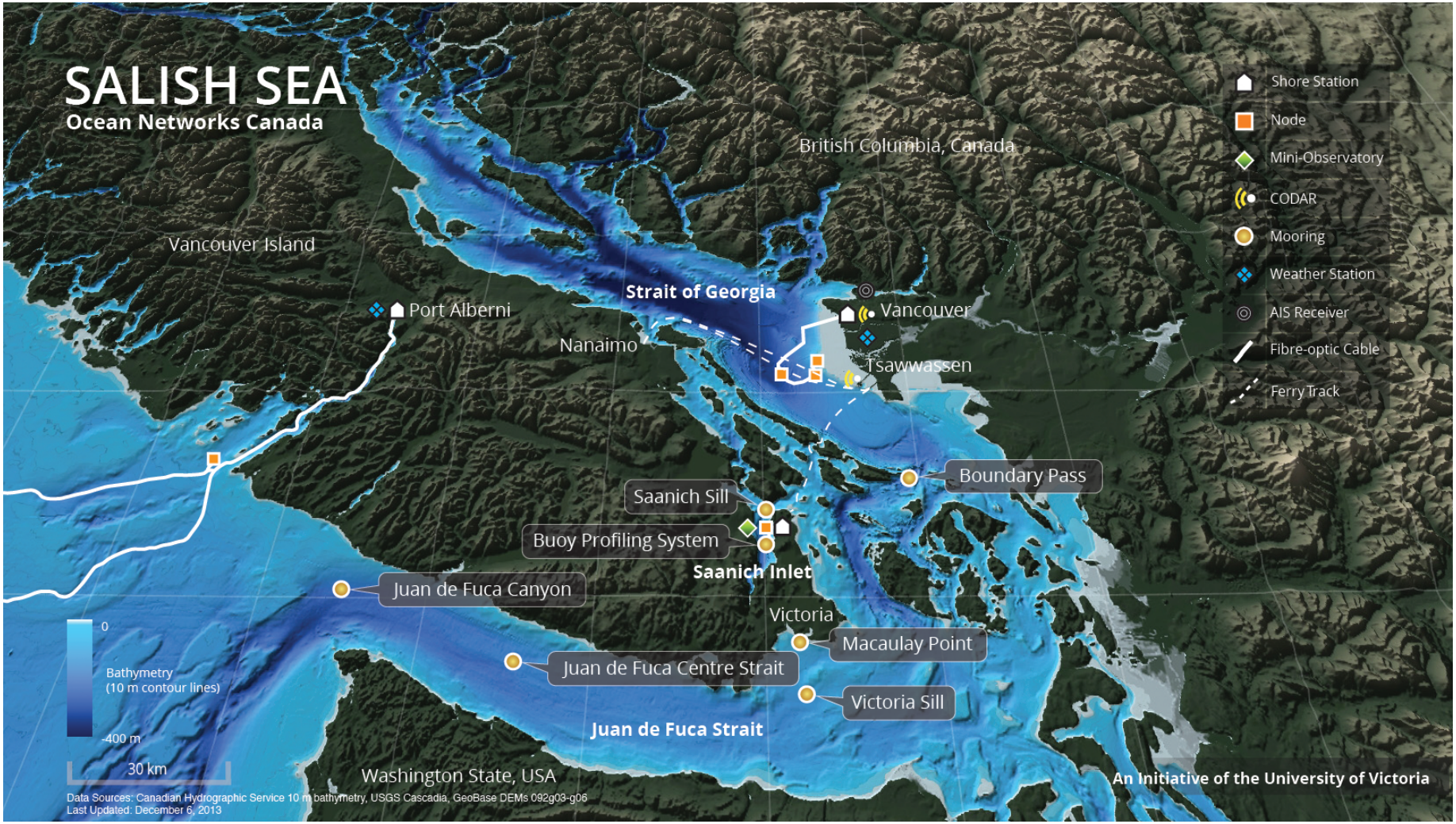
Other network contexts

Studying tram traffic in Vienna

Source: kurier.at

Other network contexts

Studying ocean flows and pollution



Source: Wikimedia

An applied problem

OVERLAPPING STOCHASTIC BLOCK MODELS WITH APPLICATION TO THE FRENCH POLITICAL BLOGOSPHERE[1]

BY PIERRE LATOUCHE, ETIENNE BIRMELÉ AND CHRISTOPHE AMBROISE

University of Evry

Complex systems in nature and in society are often represented as networks, describing the rich set of interactions between objects of interest. Many deterministic and probabilistic clustering methods have been developed to analyze such structures. Given a network, almost all of them partition the vertices into *disjoint* clusters, according to their connection profile. However, recent studies have shown that these techniques were too restrictive and that most of the existing networks contained overlapping clusters. To tackle this issue, we present in this paper the Overlapping Stochastic Block Model. Our approach allows the vertices to belong to multiple clusters, and, to some extent, generalizes the well-known Stochastic Block Model [Nowicki and Snijders (2001)]. We show that the model is generically identifiable within classes of equivalence and we propose an approximate inference procedure, based on global and local variational techniques. Using toy data sets as well as the French Political Blogosphere network and the transcriptional network of *Saccharomyces cerevisiae*, we compare our work with other approaches.

1

[1] Pierre Latouche, Etienne Birmelé, and Christophe Ambroise. "Overlapping stochastic block models with application to the french political blogosphere". In: *The Annals of Applied Statistics* (2011), pp. 309–336

An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?

An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?
- Want to classify the blogs.

An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?
- Want to classify the blogs.
- What does a good method do?

An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?
- Want to classify the blogs.
- What does a good method do?
- It produces interpretable results...

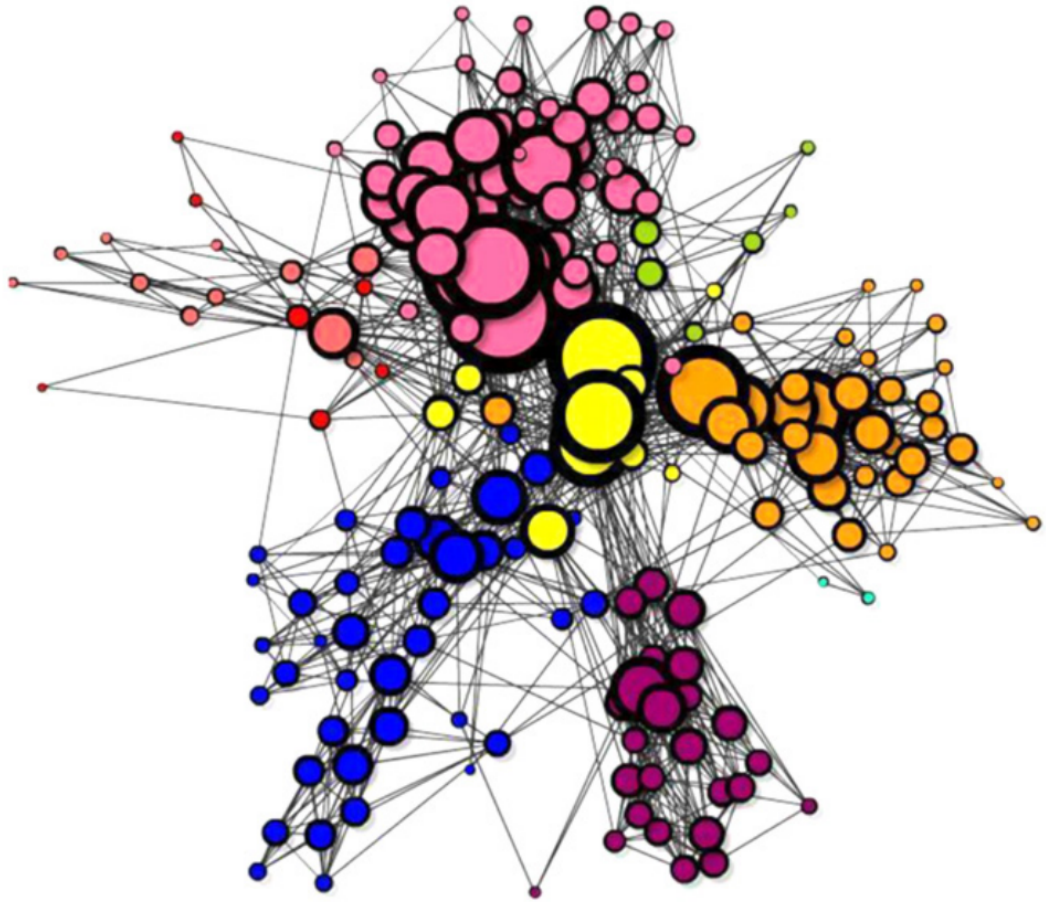
An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?
- Want to classify the blogs.
- What does a good method do?
- It produces interpretable results...
- Additional information: there are four main French political parties (UMP – republican, UDF – moderate, liberal, PS – democrat)

An applied problem

- French political blogosphere
- 196 vertices — hostnames
- 2864 edges — is there a hyperlink between two hostnames?
- Want to classify the blogs.
- What does a good method do?
- It produces interpretable results...
- Additional information: there are four main French political parties (UMP – republican, UDF – moderate, liberal, PS – democrat)
- One way to calibrate whether a method performs well is to see if it finds "subject-matter" groups.

An applied problem: the picture



[2]

Lets assume that individuals within groups are similar

[2] Hugo Zanghi, Christophe Ambroise, and Vincent Miele. "Fast online graph clustering via Erdős–Rényi mixture". In: *Pattern Recognition* 41.12 (2008), pp. 3592 –3599. ISSN: 0031-3203 – note that there are six colors...

An applied problem: some output

Stochastic Block Model

	UMP	UDF	liberal	PS	analysts	others
cluster 1	37	0	1	0	0	2
cluster 2	1	31	0	0	1	0
cluster 3	0	0	24	0	1	0
cluster 4	0	0	0	26	0	0
cluster 5	2	1	0	31	9	29

FIG. 9. *Classification of the blogs into $Q = 5$ clusters using SBM. The entry (i, j) of the matrix describes the number of blogs associated to the jth political party (column) and classified into cluster i (row). Cluster 5 corresponds to the class of outliers.*

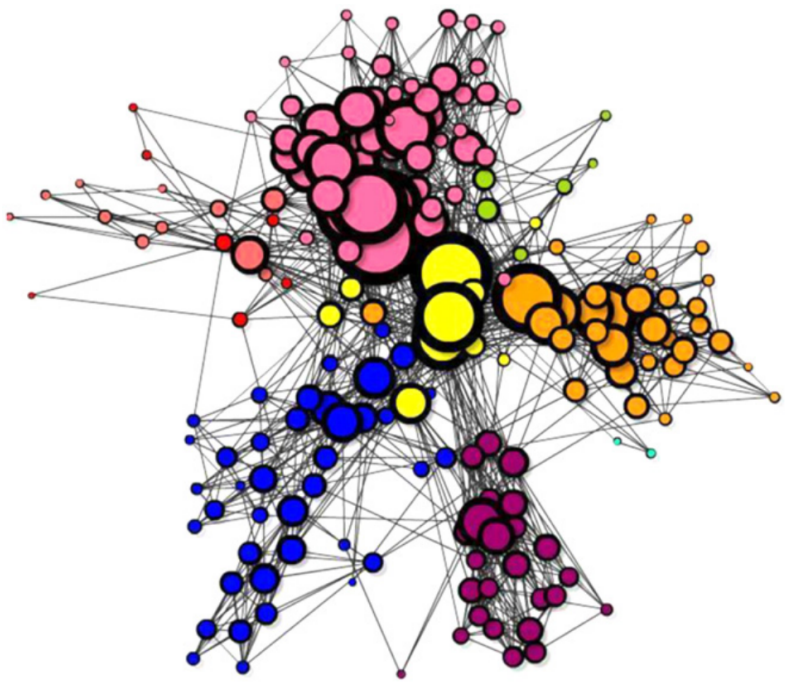
An applied problem: some output

Overlapping Stochastic Block Model

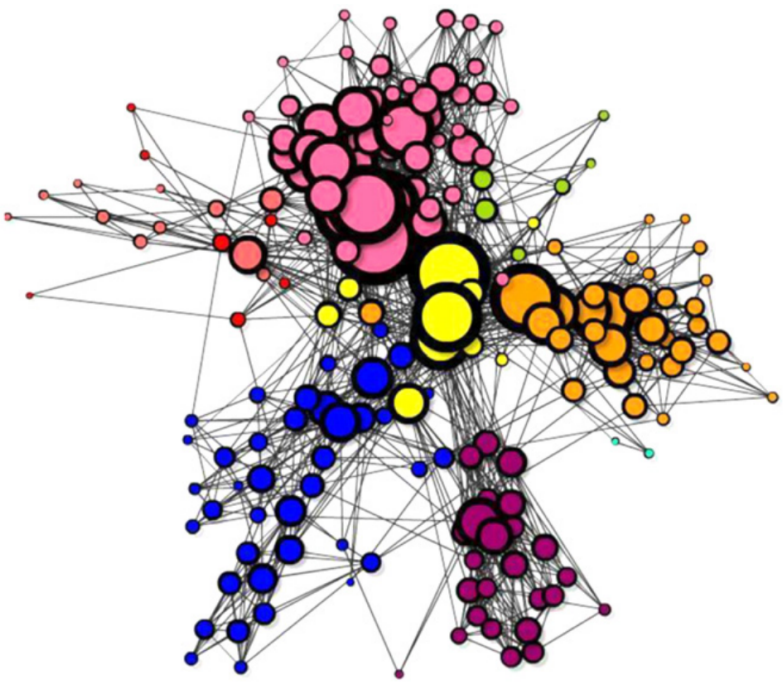
	UMP	UDF	liberal	PS	analysts	others
cluster 1	30 + 3	0 + 1	0	0	0 + 1	0
cluster 2	2 + 3	29 + 1	0	0	1 + 3	0
cluster 3	0	0	24	0	1 + 1	0
cluster 4	0	0 + 2	0	40	0 + 4	1
outliers	5	1	1	17	5	30

FIG. 7. *Classification of the blogs into $Q = 4$ clusters using OSBM. The entry (i, j) of the matrix describes the number of blogs associated to the jth political party (column) and classified into cluster i (row). Each entry distinguishes blogs which belong to a unique cluster from overlaps (single membership blogs + overlaps). The last row corresponds to the null component.*

What is actually happening here?

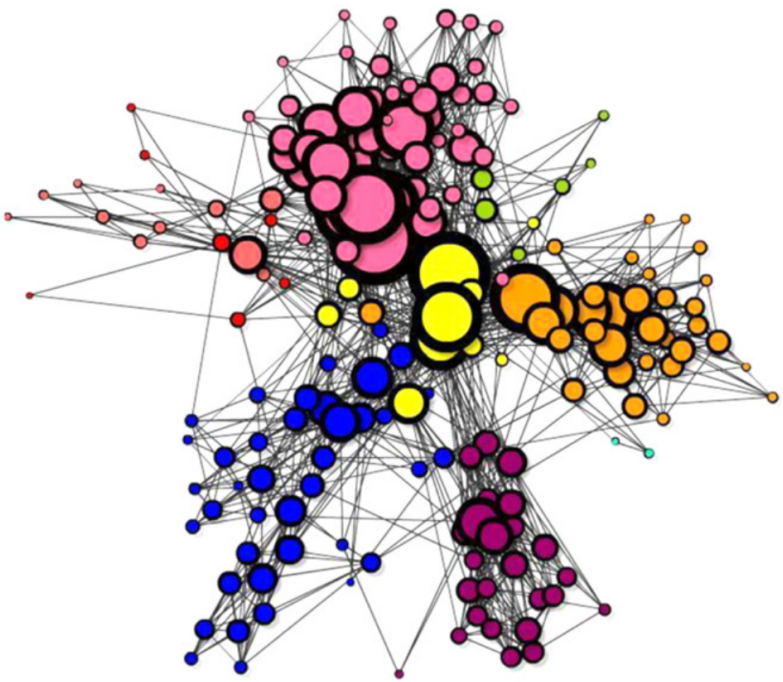


What is actually happening here?



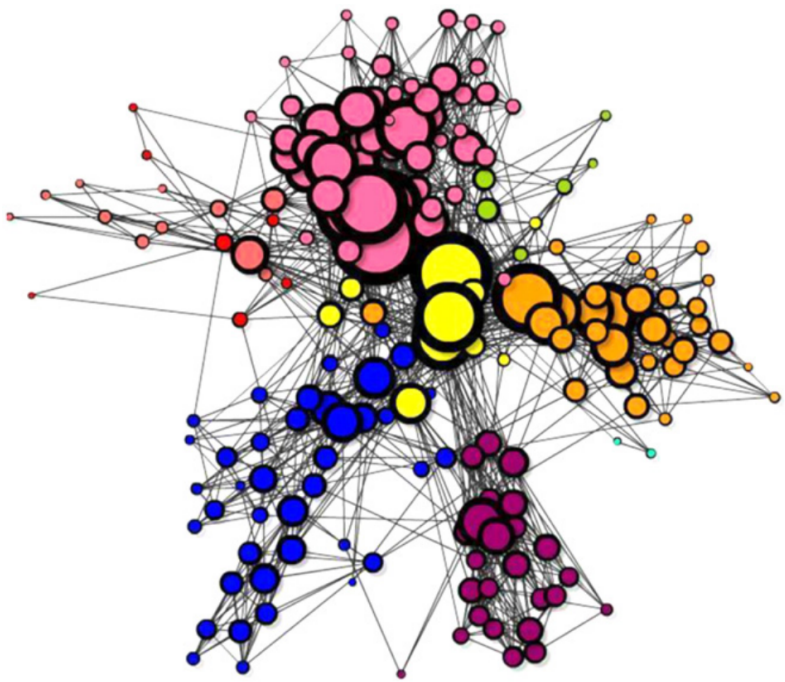
- Imagine the colors are the true groups.

What is actually happening here?



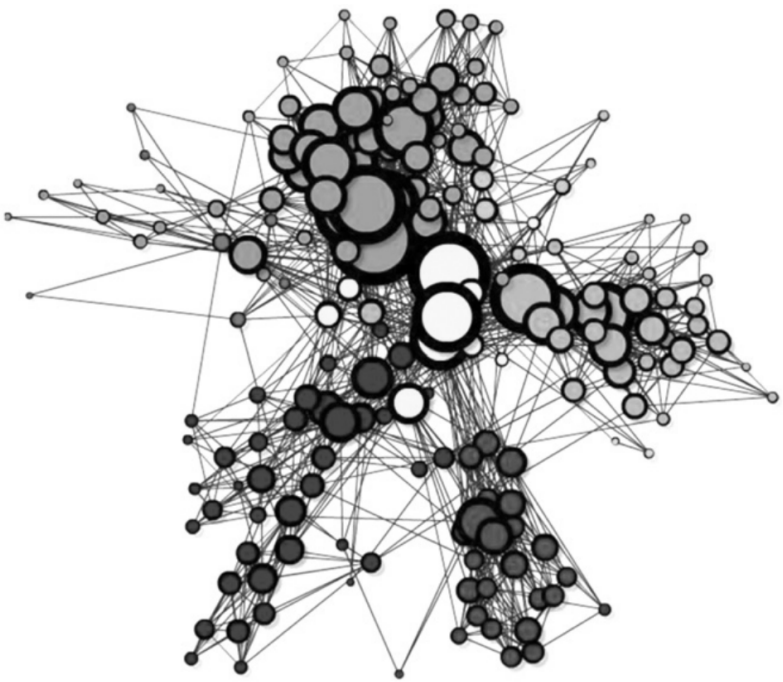
- Imagine the colors are the true groups.
- Simplest model: stochastic blockmodel — if you belong to the same group you are stochastically equivalent.

What is actually happening here?



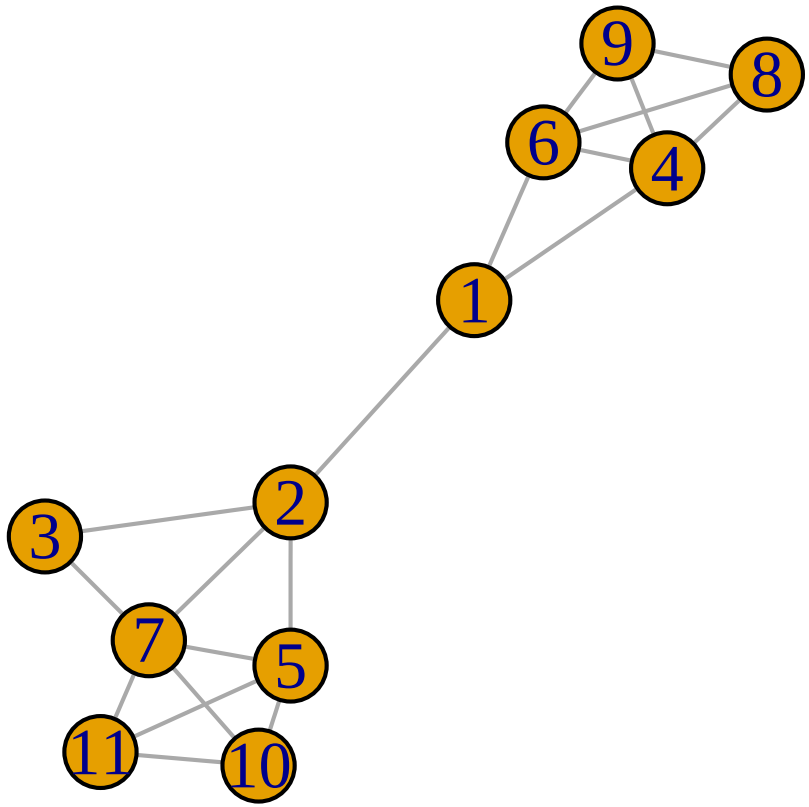
- Imagine the colors are the true groups.
- Simplest model: stochastic blockmodel — if you belong to the same group you are stochastically equivalent.
- Different methods try to find all of the stochastically equivalent nodes and put them in the same group.

What is actually happening here?



- Imagine the colors are the true groups.
- Simplest model: stochastic blockmodel — if you belong to the same group you are stochastically equivalent.
- Different methods try to find all of the stochastically equivalent nodes and put them in the same group.
- Can we do it without colors?

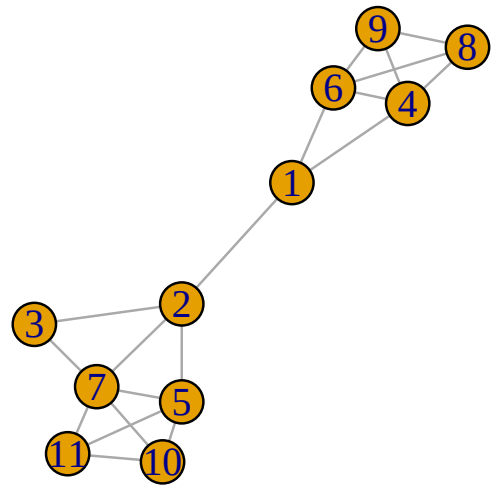
Stylized example



Stylized example — some R

We will use the igraph package extensively.

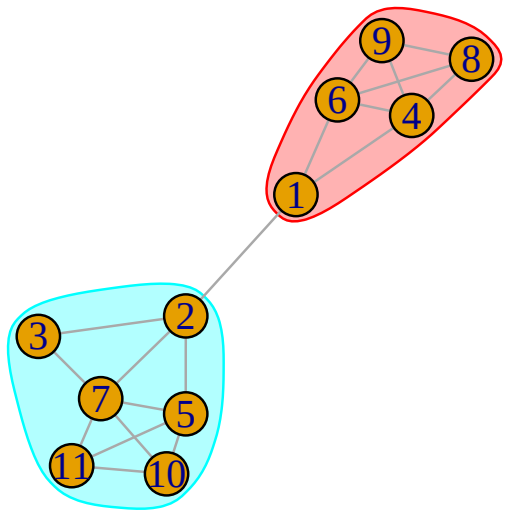
```
> membership(cluster_spinglass(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_spinglass(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_spinglass(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_optimal(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_spinglass(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_louvain(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_walktrap(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_infomap(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_fast_greedy(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_leading_eigen(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_edge_betweenness(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
```



Stylized example — some R

We will use the igraph package extensively.

```
> membership(cluster_optimal(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_spinglass(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_louvain(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_walktrap(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_infomap(first_graph))
 [1] 2 1 1 2 1 2 1 2 2 1 1
> membership(cluster_fast_greedy(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_leading_eigen(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
> membership(cluster_edge_betweenness(first_graph))
 [1] 1 2 2 1 2 1 2 1 1 2 2
```

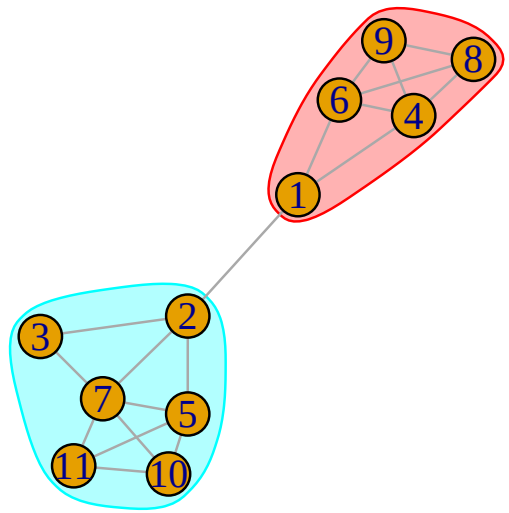


Stylized example

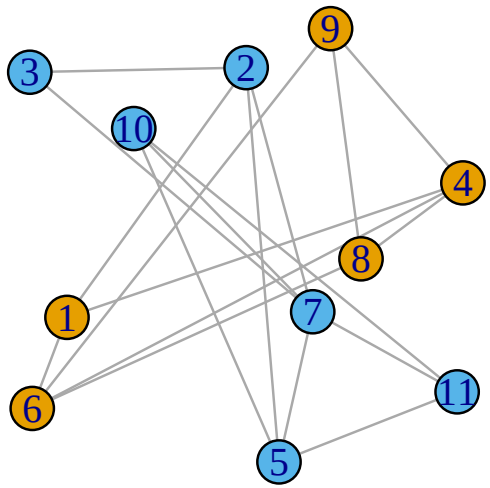
Some computer science?

- Graphs can naturally represent the flow of information between nodes.
- Famous theorems such as Max-Flow Min-Cut.
- Groups might have lots of flow inside and little flow across.

```
> min_cut(first_graph,value.only=FALSE)
$value
[1] 1
$cut
+ 1/19 edge:
[1] 1--2
$partition1
+ 6/11 vertices:
[1]  2  5  7 10 11  3
$partition2
+ 5/11 vertices:
[1] 1 4 6 8 9
```



Stylized example — troubled waters



- Graphs are never ordered nicely.
- The job of many methods is to untangle the hairball.
- This can be achieved manually, some of the tools we used above, and some basic math.

Detour through math

- What is a "mathematically" untangled graph?

Detour through math

- What is a "mathematically" untangled graph?
- Example: graph without any crossing edges.

Detour through math

- What is a "mathematically" untangled graph?
- Example: graph without any crossing edges.
- This type of graph is called a planar graph.

Detour through math

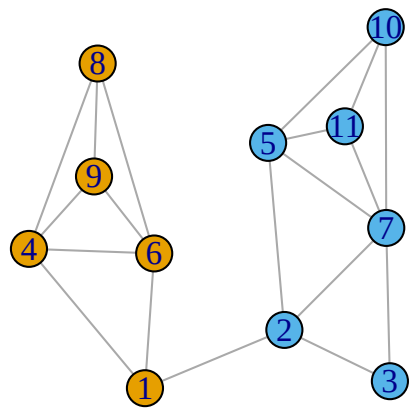
- What is a "mathematically" untangled graph?
- Example: graph without any crossing edges.
- This type of graph is called a planar graph.
- Theorem (Kuratowski): A graph is planar if and only if it does not contain a subgraph that is a subdivision of K_5 or $K_{3,3}$.

Detour through math

- What is a "mathematically" untangled graph?
- Example: graph without any crossing edges.
- This type of graph is called a planar graph.
- Theorem (Kuratowski): A graph is planar if and only if it does not contain a subgraph that is a subdivision of K_5 or $K_{3,3}$.
- Probably not that practical...

Detour through math

- What is a "mathematically" untangled graph?
- Example: graph without any crossing edges.
- This type of graph is called a planar graph.
- Theorem (Kuratowski): A graph is planar if and only if it does not contain a subgraph that is a subdivision of K_5 or $K_{3,3}$.
- Probably not that practical...



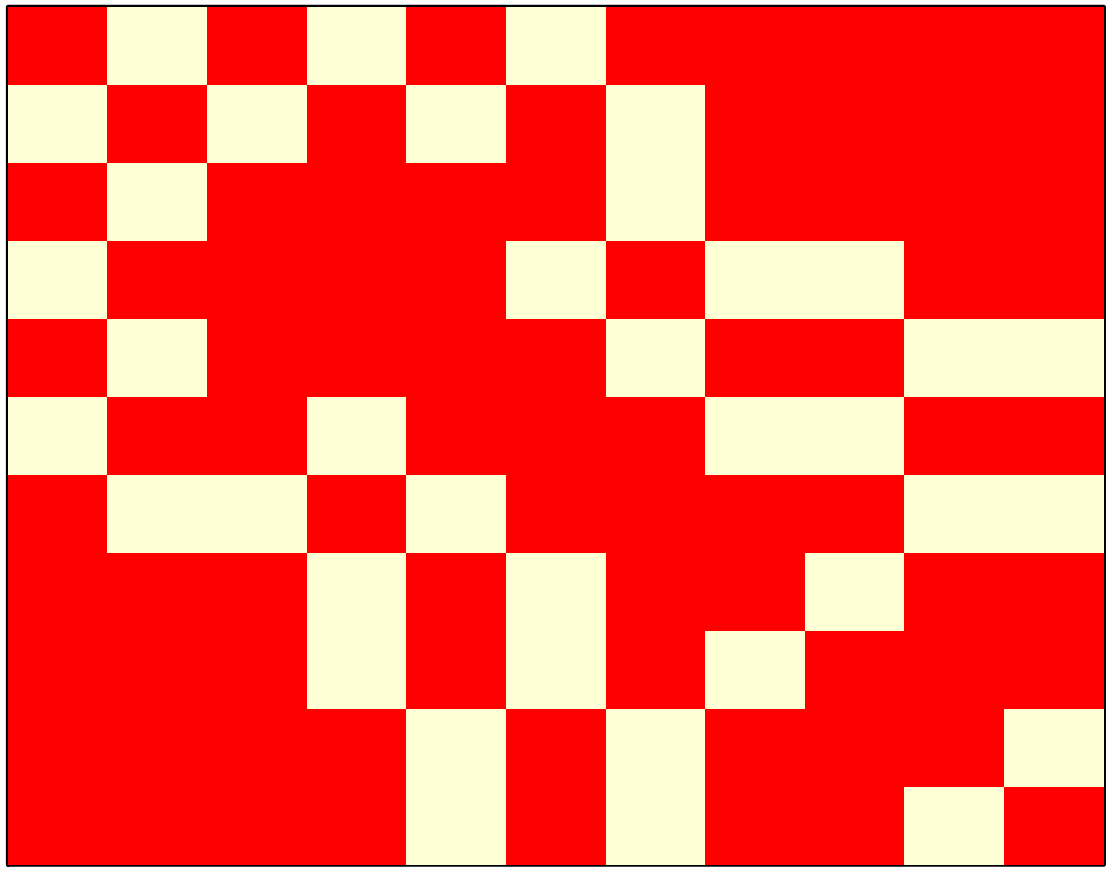
How do we really work with graphs?

- Need to represent graphs numerically.
- Lets introduce some notation. A graph G has a vertex (node) set V and an edge set E.
- If the $(i, j) \in E$ iff $(j, i) \in E$ then the graph is undirected.
- A graph can be represented by its' adjacency matrix.
- An adjacency matrix A has entries 0 and 1 where $a_{ij} = 1$ if node i is connected to node j.
- By convention $a_{ii} = 0$.

Back to the stylized example

$$A = \begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}$$

Back to the stylized example



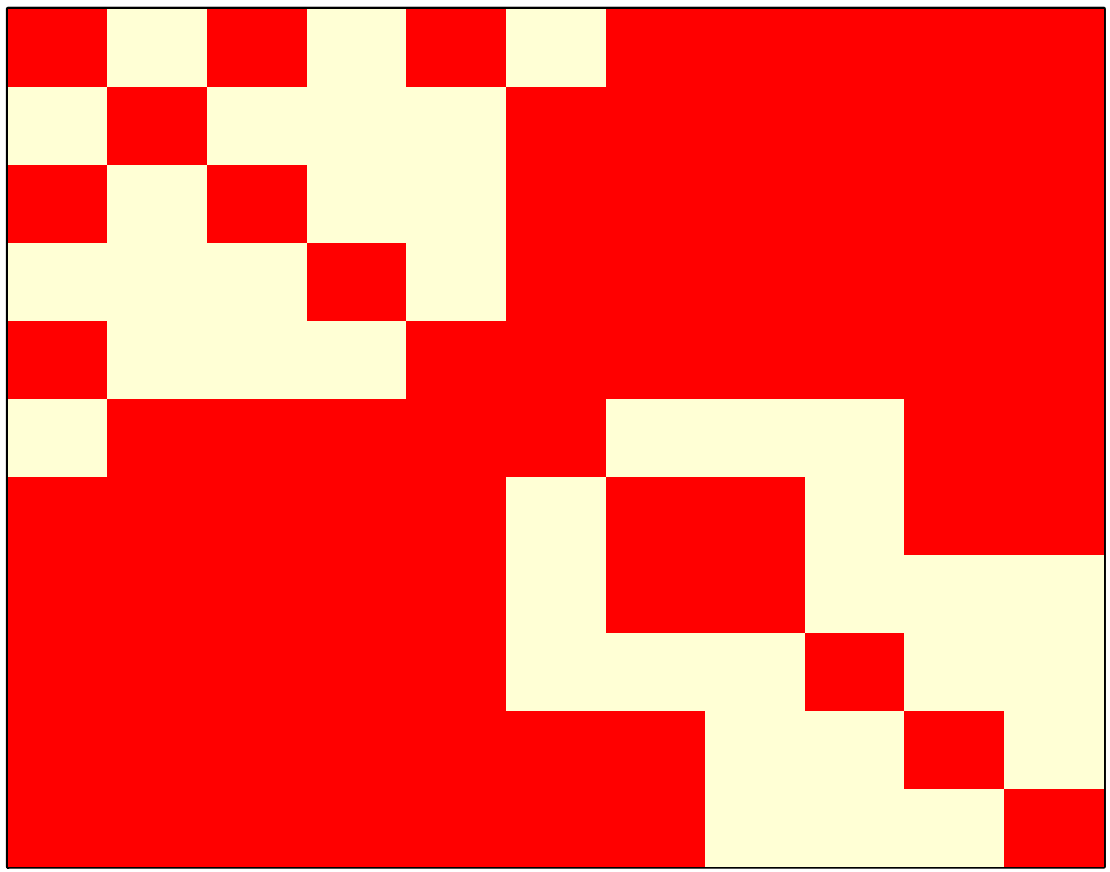
Back to the stylized example

This would be easier

$$\pi A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Back to the stylized example

This would be easier



We don't know that permutation...

- This is one of the hardest parts of the problem.
- Measure preserving transformation.
- Solution to this problem:

We don't know that permutation...

- This is one of the hardest parts of the problem.
- Measure preserving transformation.
- Solution to this problem:

a. canonical permutation and hope for the best

We don't know that permutation...

- This is one of the hardest parts of the problem.
- Measure preserving transformation.
- Solution to this problem:

a. canonical permutation and hope for the best

b. permutation agnostic methods

What are our methods for finding groups?

- Histogram methods
- Spectral methods
- Belief propagation methods
- Model-based approaches

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Chan and Airoldi (2014):

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Chan and Airoldi (2014):

- Imagine we know a good-enough permutation, call it π.

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Chan and Airoldi (2014):

- Imagine we know a good-enough permutation, call it π.
- Transform the adjacency matrix: $A \rightarrow \pi A$

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Chan and Airoldi (2014):

- Imagine we know a good-enough permutation, call it π.
- Transform the adjacency matrix: $A \rightarrow \pi A$
- "Smooth" the transformed adjacency matrix.

Histogram Methods

Several approaches

Airoldi, Costa and Chan (2013):

- Compute some distance measure between nodes in a graph.
- Cluster nodes based on this distance measure.
- Estimate the block probabilities based on the assigned nodes.

Chan and Airoldi (2014):

- Imagine we know a good-enough permutation, call it π.
- Transform the adjacency matrix: $A \to \pi A$
- "Smooth" the transformed adjacency matrix.
- Minimize distance to a desirable object (such as a smooth function or a piecewise constant function)

Spectral Methods

Lots of theory developed in Rohe and Yu (2011), Rohe, Chatterjee, and Yu (2011)

Spectral Methods

Lots of theory developed in Rohe and Yu (2011), Rohe, Chatterjee, and Yu (2011)

- Define an object of interest: For example, the graph Laplacian $L = I - D^{-1/2} A D^{-1/2}$

Spectral Methods

Lots of theory developed in Rohe and Yu (2011), Rohe, Chatterjee, and Yu (2011)

- Define an object of interest: For example, the graph Laplacian $L = I - D^{-1/2} A D^{-1/2}$
- Find the eigenvectors associated with its largest k eigenvalues (say in absolute value)

Spectral Methods

Lots of theory developed in Rohe and Yu (2011), Rohe, Chatterjee, and Yu (2011)

- Define an object of interest: For example, the graph Laplacian $L = I - D^{-1/2} A D^{-1/2}$
- Find the eigenvectors associated with its largest k eigenvalues (say in absolute value)
- Cluster the rows of the $k \times n$ eigenvector matrix into k clusters.

Spectral Methods

Lots of theory developed in Rohe and Yu (2011), Rohe, Chatterjee, and Yu (2011)

- Define an object of interest: For example, the graph Laplacian $L = I - D^{-1/2} A D^{-1/2}$
- Find the eigenvectors associated with its largest k eigenvalues (say in absolute value)
- Cluster the rows of the $k \times n$ eigenvector matrix into k clusters.
- If need be, estimate the probabilities within each cluster/block.

Belief propagation methods

- A. Decelle, *F. Krzakala*, C. Moore, and *L. Zdeborova*, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, Phys. Rev. E 84 (2011), 066106.
- Essentially start with some group assignment for each node, broadcast to nearby nodes and update.
- Loads of recent work on (theoretical) optimality of these methods.

Model based approaches

The stochastic blockmodel has a natural data generative form:

Model based approaches

The stochastic blockmodel has a natural data generative form:

- Let $z_1, \ldots, z_n$ be the block memberships of n nodes

Model based approaches

The stochastic blockmodel has a natural data generative form:

- Let $z_1, \ldots, z_n$ be the block memberships of n nodes
- Let B be the matrix of probabilities of connections between blocks.

Model based approaches

The stochastic blockmodel has a natural data generative form:

- Let $z_1, \ldots, z_n$ be the block memberships of n nodes
- Let B be the matrix of probabilities of connections between blocks.
- There is an edge between nodes i and j with probability $B_{b_i b_j}$.

Model based approaches

The stochastic blockmodel has a natural data generative form:

- Let $z_1, \ldots, z_n$ be the block memberships of n nodes
- Let B be the matrix of probabilities of connections between blocks.
- There is an edge between nodes i and j with probability $B_{b_i b_j}$.

  By specifying a prior for the block memberships and observing an adjacency matrix A we have all the ingredients to estimate B and block membership.

What do we have to look forward to

- Thursday: probability
- Tuesday: probability and karate
- Lab will start next week.