

# Structure and dynamics of complex networks

April 23, 2020

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

# Motifs and communities

April 23, 2020

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

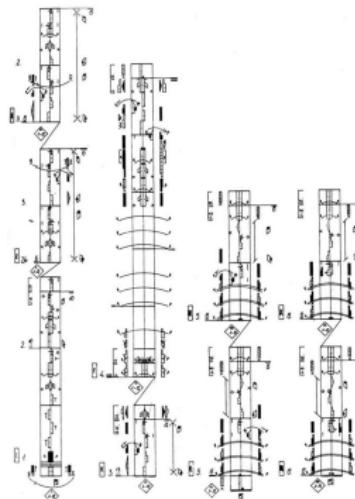
Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



## NETWORK MOTIFS

# Motifs and communities: the mesoscopic scale

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

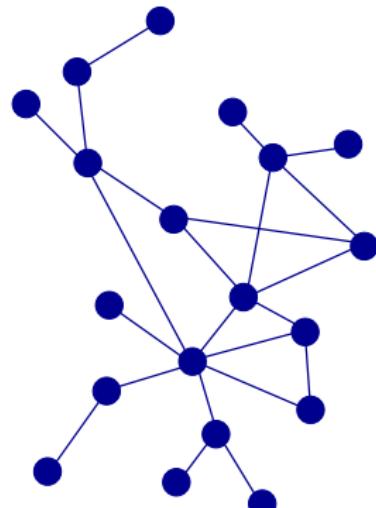
Modularity

Overlapping communities

The Clique

Percolation Method

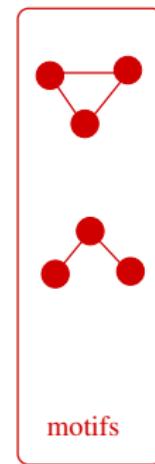
The Infomap method



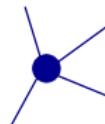
whole network:  $p(k), \langle l \rangle$



communities



motifs



individual nodes:  $k_i c_i$

# Network Motifs

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Network motif

- A motif is a small, connected subgraph in which the configuration of the links is predefined.
- E.g., for undirected networks we have only 2 types of triadic motifs:



- In the directed case, the number of motifs is higher, e.g., already 13 triadic motifs.



# Network Motifs

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Network motif

- A motif is a small, connected subgraph in which the configuration of the links is predefined.
- E.g., for undirected networks we have only 2 types of triadic motifs:



- In the directed case, the number of motifs is higher, e.g., already 13 triadic motifs.



# Network Motifs

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Network motif

- A motif is a small, connected subgraph in which the configuration of the links is predefined.
- E.g., for undirected networks we have only 2 types of triadic motifs:



- In the directed case, the number of motifs is higher, e.g., already 13 triadic motifs.



# Network Motifs

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Network motif

- A motif is a small, connected subgraph in which the configuration of the links is predefined.
- E.g., for undirected networks we have only 2 types of triadic motifs:



- In the directed case, the number of motifs is higher, e.g., already 13 triadic motifs.



# Network Motifs

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Network motif

- A motif is a small, connected subgraph in which the configuration of the links is predefined.
- E.g., for undirected networks we have only 2 types of triadic motifs:



- In the directed case, the number of motifs is higher, e.g., already 13 triadic motifs.



# Network Motifs

Why are they important?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

## Why are network motifs important?

- In some cases they have a likely role in, e.g., information processing or controlling.
- In many cases their statistics show interesting features:
  - E.g., in a given network, one motif may occur very often, whereas another motif may be almost completely absent.
  - When comparing the frequencies of a given motif between different networks, again, we may observe large differences, e.g., the motif can be very frequent in some of the networks, and quite rare in the other networks.

# Feed Forward Loop

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

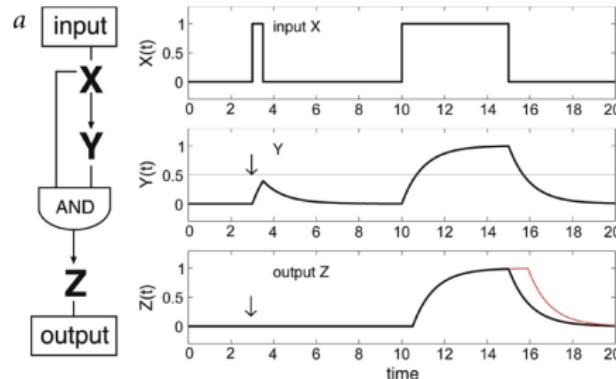
The Clique

Percolation Method

The Infomap method

Example for likely role of motifs:

- It is quite plausible that the Feed Forward Loop (FFL) can serve as a sort noise filter in biological networks:



# $z$ -score

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

How to quantify whether a motif can be considered as very frequent or very rare?

# $z$ -SCORE

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

How to quantify whether a motif can be considered as very frequent or very rare?

- We randomise the network keeping the degree sequence, using the concept of link-randomisation or node-randomisation (that were discussed in Lecture 8.).

# $z$ -SCORE

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

How to quantify whether a motif can be considered as very frequent or very rare?

- We randomise the network keeping the degree sequence, using the concept of link-randomisation or node-randomisation (that were discussed in Lecture 8.).
- and compare the number of occurrences in the original network with the number of occurrences in the randomised one:
  - if there are significantly more of the given motif in the original network, we can say that it is "over expressed",
  - if there are significantly less of the given motif in the original network, we can say that it is "under expressed".

# $z$ -SCORE

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

## Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- OK, but what does it mean to observe a significant difference between the number of occurrences?

# $z$ -SCORE

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- OK, but what does it mean to observe a significant difference between the number of occurrences?
- We can decide based on the  $z$ -score:

$$z(m_i) \equiv \frac{\langle m_i \rangle_{\text{real}} - \langle m_i \rangle_{\text{rand}}}{\sigma_{\text{rand}}},$$

where  $\langle m_i \rangle_{\text{real}}$  is the number of motif  $m_i$  in the original network,  $\langle m_i \rangle_{\text{rand}}$  is the average number of the same motif in a randomised network, and  $\sigma_{\text{rand}}$  is corresponding to the standard deviation of the number of the motif in randomised networks.

# $z$ -SCORE

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- OK, but what does it mean to observe a significant difference between the number of occurrences?
- We can decide based on the  $z$ -score:

$$z(m_i) \equiv \frac{\langle m_i \rangle_{\text{real}} - \langle m_i \rangle_{\text{rand}}}{\sigma_{\text{rand}}},$$

where  $\langle m_i \rangle_{\text{real}}$  is the number of motif  $m_i$  in the original network,  $\langle m_i \rangle_{\text{rand}}$  is the average number of the same motif in a randomised network, and  $\sigma_{\text{rand}}$  is corresponding to the standard deviation of the number of the motif in randomised networks.

If the  $z$ -score is larger than 1, or smaller than -1, the difference between the original and the randomised is strong, whereas for  $z$ -scores with an absolute value larger than 3, the difference is significant by all means.

# $z$ -score

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

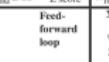
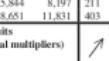
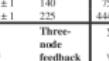
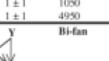
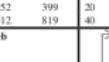
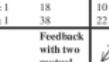
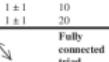
Overlapping communities

The Clique

Percolation Method

The Infomap method

Top  $z$ -scores and corresponding motifs measured in a couple of networks:

Network	Nodes	Edges	$N_{real}$	$N_{rand} \pm SD$	Z score	$N_{real}$	$N_{rand} \pm SD$	Z score	$N_{real}$	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)											
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> <sup>+</sup>	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons											
<i>C. elegans</i> <sup>†</sup>	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs											
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Yihua	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Couchella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skewwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)											
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)											
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s385 <sup>‡</sup>	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web											
nd.edu <sup>§</sup>	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

# Motif significance profile

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- Based on  $z$ -scores measured for the different motifs, we can prepare the **motif significance profile**, which is like a finger print of the network.
- The value of the significance profile for a given motif  $i$  is defined as the "normalised"  $z$ -score, written as

$$SP_i \equiv \frac{z_i}{\sqrt{\sum_j z_j^2}},$$

where the summation in the denominator runs over all possible motifs (e.g., the 13 motifs in case of a triadic significance profile).

# Motif significance profile

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

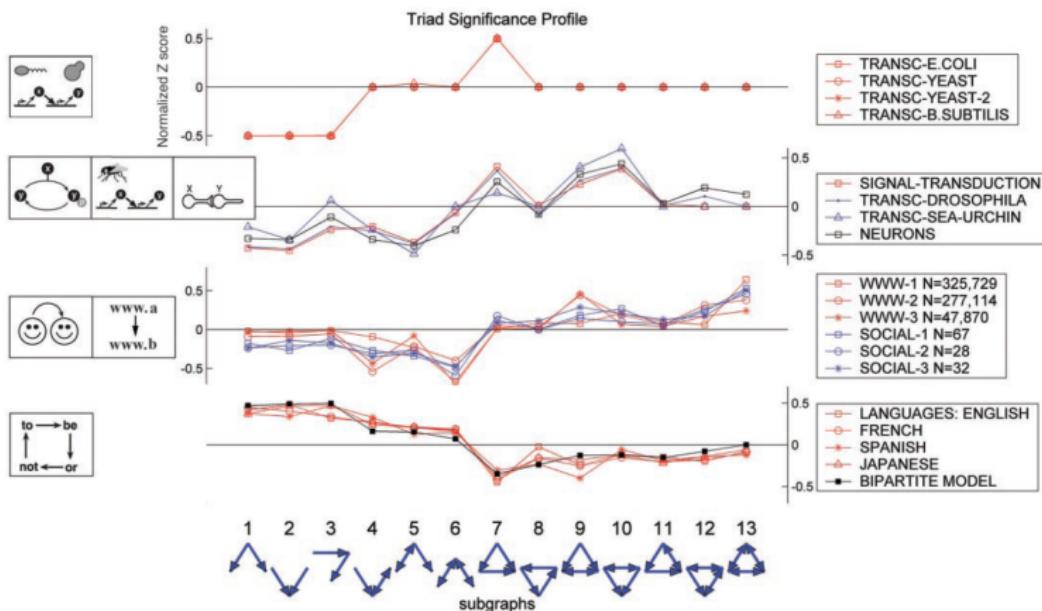
Overlapping communities

The Clique

Percolation Method

The Infomap method

The triadic motif significance profile for a couple of networks, forming 4 "families":

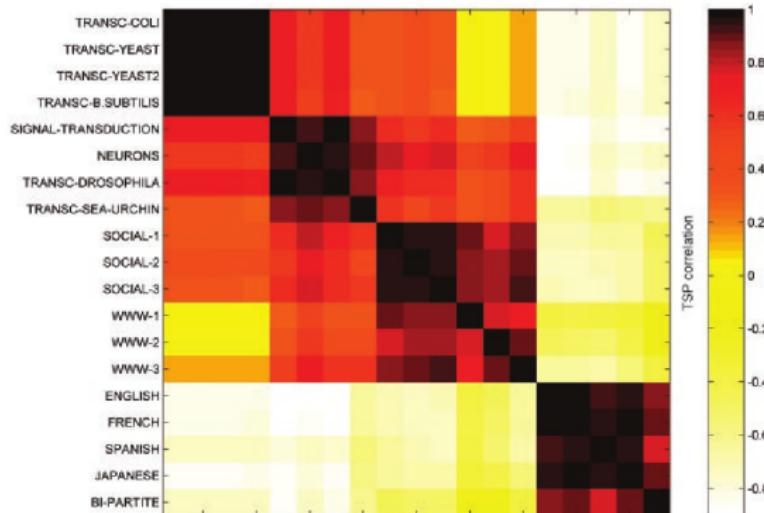


## Motif significance profile

## Motifs and communities

## Motif significance profile

## Correlation between significance profiles:



# Motifs summary

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- Motifs correspond to small sub-graphs with a predefined configuration of the links.
- In real networks the frequencies of the different motifs can differ significantly from what we would expect in a random network with the same degree distribution.
- By constructing the motif significance profile we obtain a sort of fingerprint, based on which we can
  - compare the profile value of the different motifs with each other to see e.g., which motif is the most over expressed or under expressed,
  - compare different networks with each other.

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# COMMUNITIES

# Motifs and communities: the mesoscopic scale

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

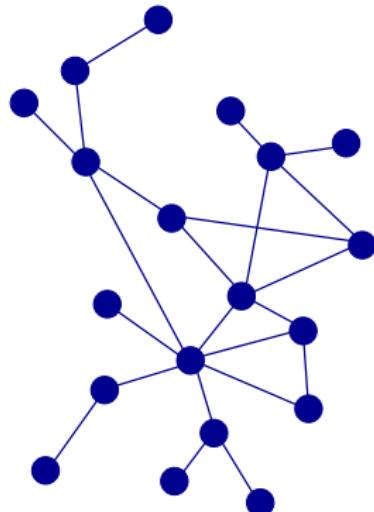
Modularity

Overlapping communities

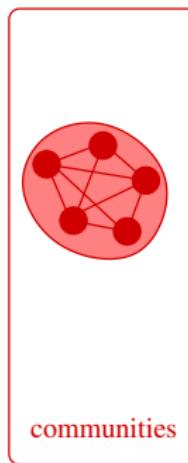
The Clique

Percolation Method

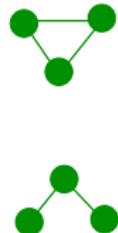
The Infomap method



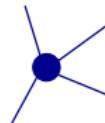
whole network:  $p(k), \langle l \rangle$



communities



motifs



individual nodes:  $k_i$   $c_i$

# What are communities?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

**Communities, modules, clusters, or cohesive groups:**  
more highly interconnected parts in networks with no widely accepted  
unique definition.

## Examples

- A family, or a group of friends in a social network.
- A group of proteins having the same function or taking part in the same process in a protein interaction graph.
- Interlinked Web pages with the same topic.

• ...

# What are communities?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

**Communities, modules, clusters, or cohesive groups:**  
more highly interconnected parts in networks with no widely accepted  
unique definition.

## Examples

- A family, or a group of friends in a social network.
- A group of proteins having the same function or taking part in the same process in a protein interaction graph.
- Interlinked Web pages with the same topic.
- ...

# What are communities?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

**Communities, modules, clusters, or cohesive groups:**  
more highly interconnected parts in networks with no widely accepted  
unique definition.

## Examples

- A family, or a group of friends in a social network.
- A group of proteins having the same function or taking part in the same process in a protein interaction graph.
- Interlinked Web pages with the same topic.
- ...

# What are communities?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique  
Percolation Method

The Infomap method

**Communities, modules, clusters, or cohesive groups:**  
more highly interconnected parts in networks with no widely accepted unique definition.

## Examples

- A family, or a group of friends in a social network.
- A group of proteins having the same function or taking part in the same process in a protein interaction graph.
- Interlinked Web pages with the same topic.

• • •

# What are communities?

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique  
Percolation Method

The Infomap method

**Communities, modules, clusters, or cohesive groups:**  
more highly interconnected parts in networks with no widely accepted unique definition.

## Examples

- A family, or a group of friends in a social network.
- A group of proteins having the same function or taking part in the same process in a protein interaction graph.
- Interlinked Web pages with the same topic.
- ...

# Communities

## Examples

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

##### Examples

Hierarchical clustering

The Girvan-Newman algorithm

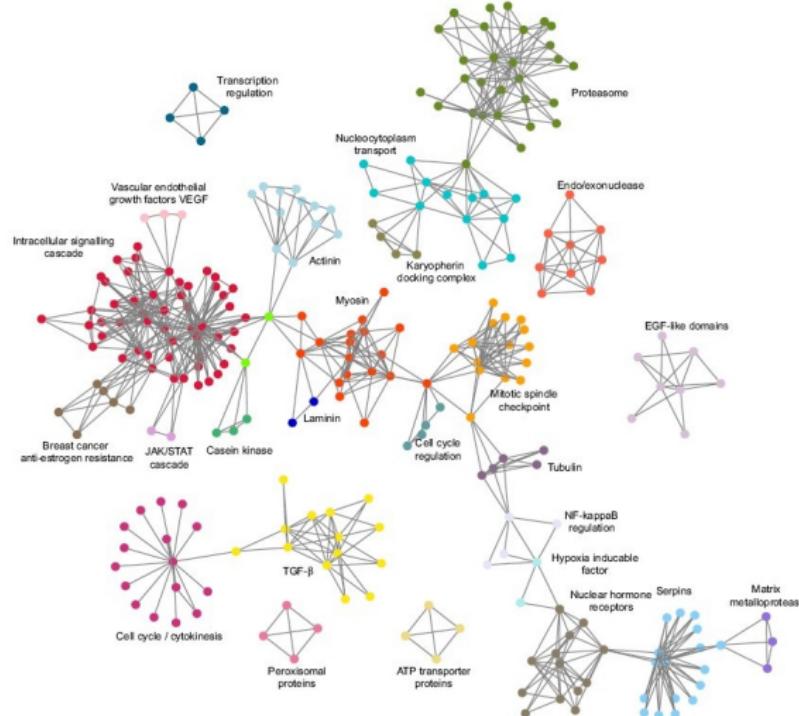
Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# Communities

## Examples

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

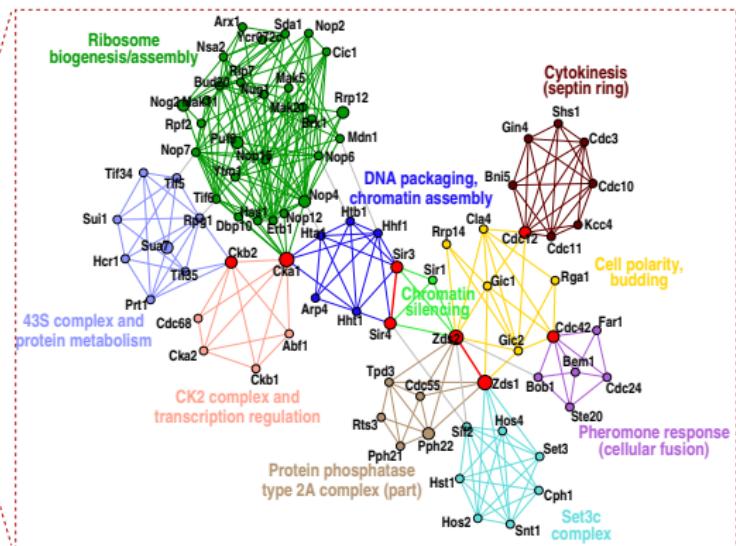
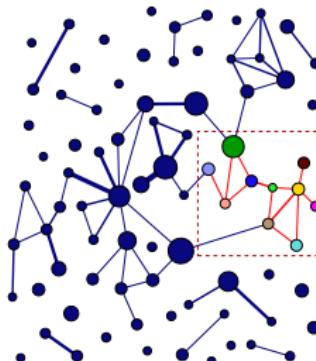
Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# Communities

## Examples

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile  
profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

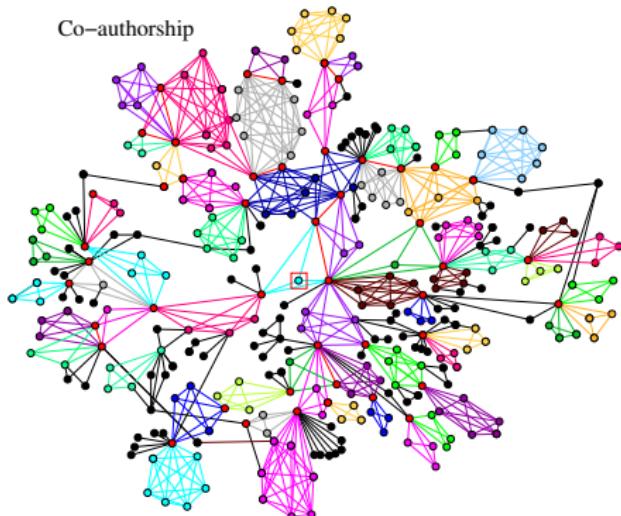
Overlapping communities

The Clique

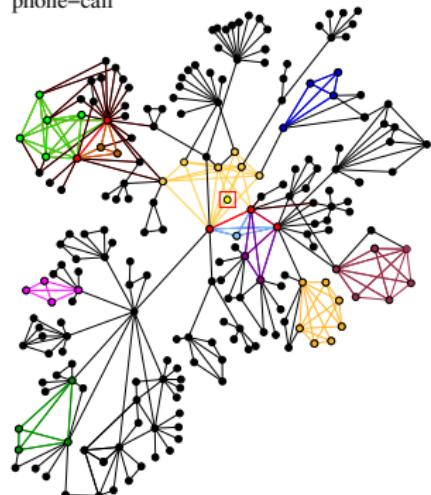
Percolation Method

The Infomap method

Co-authorship



phone-call



# Communities

## Examples

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

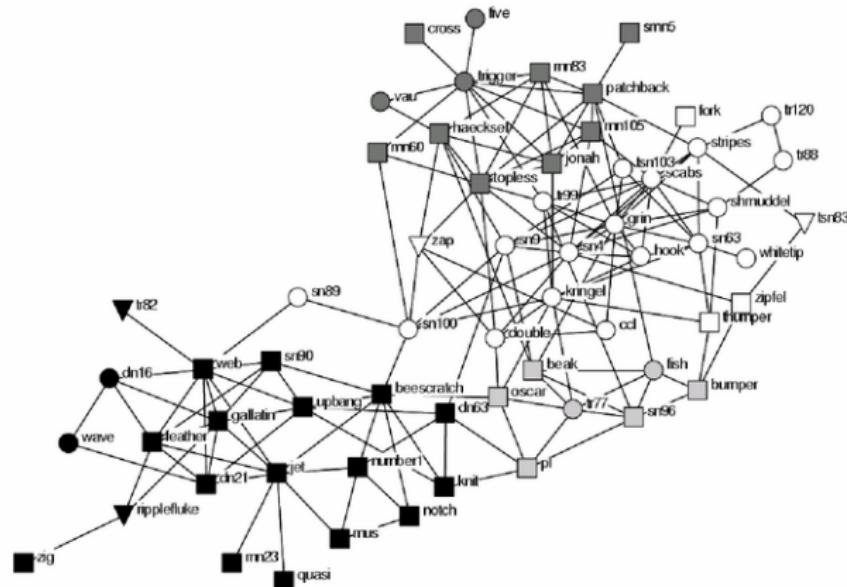
Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# Importance of communities

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## Why are communities important?

# Importance of communities

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

### Community finding is important...

- data clustering,
- recommendation systems,
- a higher level of organisation,
- fast spreading inside communities,
- prediction of protein functions

# Diversity of community finding methods

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

## Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

There are at least 40-50 community finding methods on the market...  
Which one to choose?

- global methods ↔ local methods,
- number of parameters?
- size of the system ↔ complexity of the method,
- overlapping communities ↔ isolated communities.

For a detailed review of the different ideas and methods about community finding see these excellent review papers:  
S. Fortunato, *Physics Reports* **486**, 75-174 (2010)  
S. Fortunato and D. Hric, *Physics Reports* **659**, 1-44 (2016)

# Hierarchical clustering

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- One of the most widely known data clustering methods in general is **hierarchical clustering**.

# Hierarchical clustering

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- One of the most widely known data clustering methods in general is **hierarchical clustering**.
- Thus, if we would like to find communities (or clusters) in networks, why can't we use this approach?

# Hierarchical clustering

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- One of the most widely known data clustering methods in general is **hierarchical clustering**.
- Thus, if we would like to find communities (or clusters) in networks, why can't we use this approach?
- Although hierarchical clustering was shown to be successful in many cases, it has some drawbacks when applied to networks.

# Hierarchical clustering

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- One of the most widely known data clustering methods in general is **hierarchical clustering**.
- Thus, if we would like to find communities (or clusters) in networks, why can't we use this approach?
- Although hierarchical clustering was shown to be successful in many cases, it has some drawbacks when applied to networks.
- In the following we shortly overview the main concepts of hierarchical clustering, and the problems we run into when trying to use it on networks.

# Hierarchical clustering in general

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

## Illustration:



# Hierarchical clustering

## in general

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

### Illustration:



# Hierarchical clustering in general

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

## Illustration:



# Hierarchical clustering in general

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

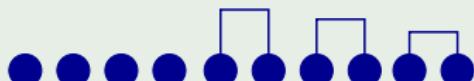
The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

## Illustration:



# Hierarchical clustering in general

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

## Illustration:



# Hierarchical clustering

## in general

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

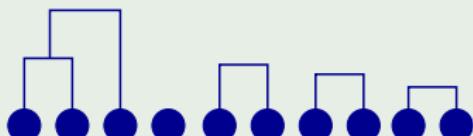
The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

### Illustration:



# Hierarchical clustering

## in general

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

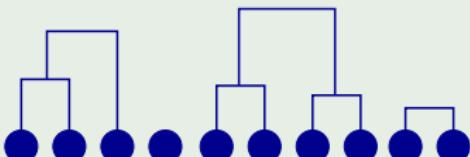
The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

### Illustration:



# Hierarchical clustering

## in general

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

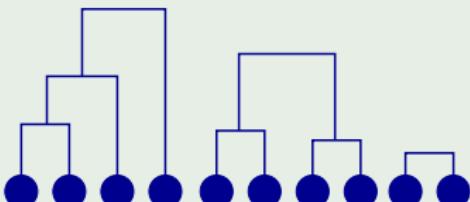
The Clique

Percolation Method

The Infomap method

- We define a **similarity** (or distance) between data points,
- and agglomerate the data points in an iterative manner, building up a **dendrogram**.
- At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

### Illustration:



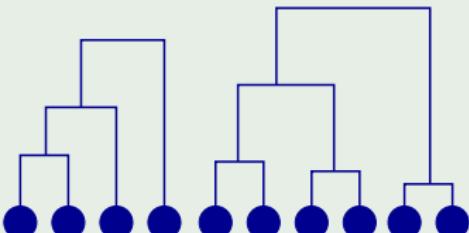
## Hierarchical clustering in general

## Motifs and communities

Hierarchical clustering

- We define a **similarity** (or distance) between data points,
  - and agglomerate the data points in an iterative manner, building up a **dendrogram**.
  - At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

### Illustration:



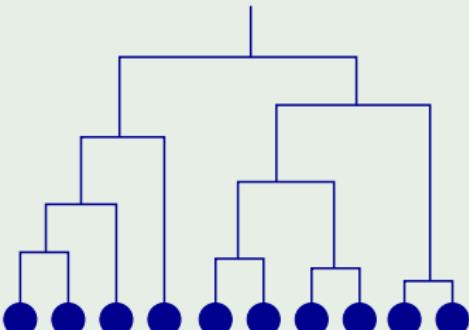
## Hierarchical clustering in general

## Motifs and communities

Hierarchical clustering

- We define a **similarity** (or distance) between data points,
  - and agglomerate the data points in an iterative manner, building up a **dendrogram**.
  - At any step, the two points (or clusters) to be merged are the ones which are the most similar (or are the closest to each other).

## Illustration:



# Hierarchical clustering in networks

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

### Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## How to apply this framework to networks?

- data points → nodes,
- similarity?

# Hierarchical clustering in networks

Motifs and  
communities

Motifs

Definition

$z$ -score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

## How to apply this framework to networks?

- data points → nodes,
- similarity (distance):
  - number of independent paths,

# Hierarchical clustering in networks

Motifs and  
communities

Motifs

Definition

$z$ -score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

## How to apply this framework to networks?

- data points → nodes,
- similarity (distance):
  - number of independent paths,
  - row (column) similarity in the adjacency matrix:

$$s_{ij} = \sqrt{\sum_k (A_{ik} - A_{jk})^2}$$

# Hierarchical clustering in networks

Motifs and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

The Clique

Percolation Method

The Infomap method

## How to apply this framework to networks?

- data points → nodes,
- similarity (distance):
  - number of independent paths,
  - row (column) similarity in the adjacency matrix:

$$s_{ij} = \sqrt{\sum_k (A_{ik} - A_{jk})^2}$$

- Pearson-correlation:

$$s_{ij} = \frac{\sum_k (A_{ik} - \mu_i)(A_{jk} - \mu_j)}{N\sigma_i\sigma_j},$$

$$\mu_i = \frac{1}{N} \sum_k A_{ik}, \quad \sigma_i = \sqrt{\sum_k (A_{ik} - \mu_i)^2}$$

# Hierarchical clustering

## Drawback I.

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile  
profile

#### Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

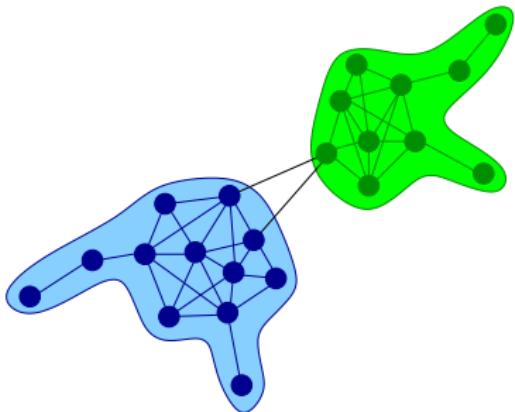
The Clique

Percolation Method

The Infomap method

- The agglomeration of the dense cores precedes the attachment of the weakly connected parts.

Illustration:



# Hierarchical clustering

## Drawback I.

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile  
profile

#### Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

Overlapping  
communities

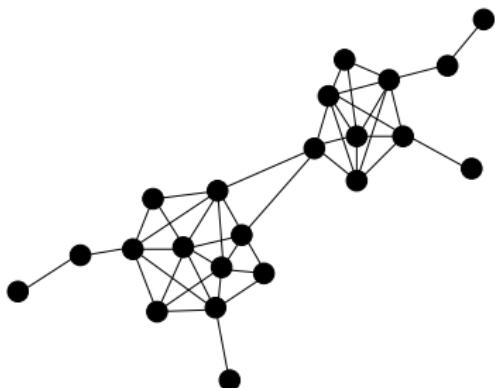
The Clique

Percolation Method

The Infomap method

- The agglomeration of the dense cores precedes the attachment of the weakly connected parts.

Illustration:



# Hierarchical clustering

## Drawback I.

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

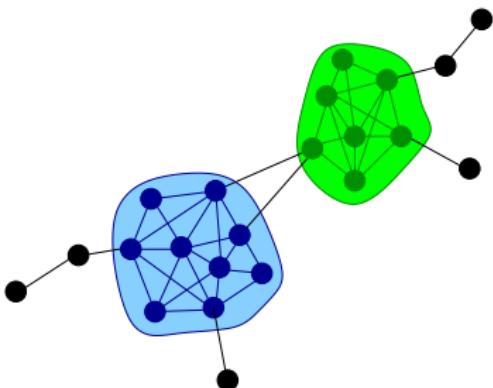
The Clique

Percolation Method

The Infomap method

- The agglomeration of the dense cores precedes the attachment of the weakly connected parts.

### Illustration:



# Hierarchical clustering

## Drawback I.

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

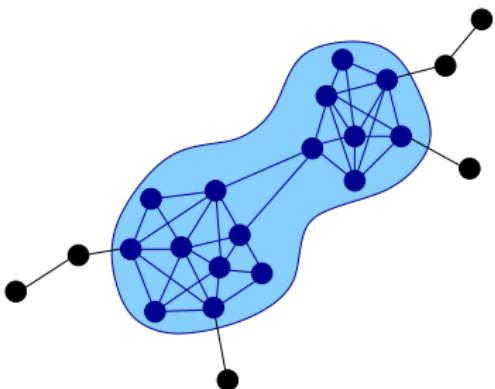
The Clique

Percolation Method

The Infomap method

- The agglomeration of the dense cores precedes the attachment of the weakly connected parts.

Illustration:



# Hierarchical clustering

## Drawback II.

### Motifs and communities

#### Motifs

Definition

z-score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

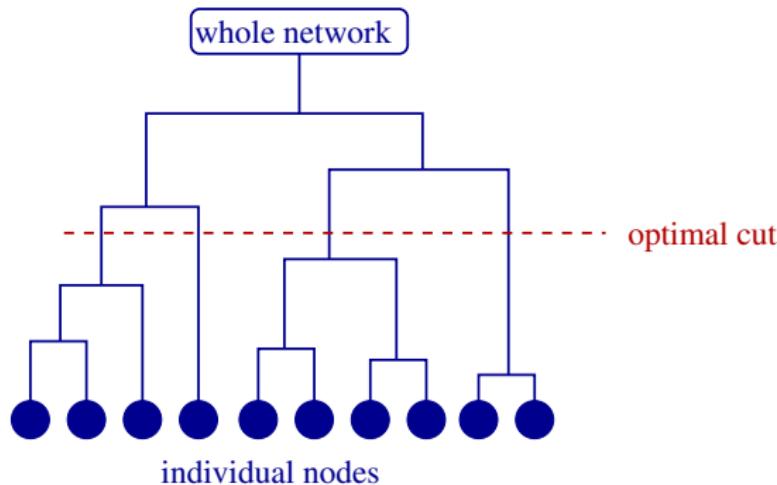
Overlapping communities

The Clique

Percolation Method

The Infomap method

- Where to cut the dendrogram?



- How can we tell which is the optimal cut for obtaining the best partitioning of the network into communities in case of a large network? The algorithm of hierarchical clustering gives no clue what so ever about that...

# The Girvan-Newman algorithm

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

One of first and most famous community finding algorithms for networks was proposed by M. E. J. Newman and M. Girvan, and was intended to cure both of the problems found when trying to partition a network into modules using the idea of hierarchical clustering.

# The Girvan-Newman algorithm

## Motivation

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

#### The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

How to avoid the early agglomeration of the cores in hierarchical clustering?

# The Girvan-Newman algorithm

## Motivation

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

How to avoid the early agglomeration of the cores in hierarchical clustering?

→ Using a **divisive** method concentrating on the **links between communities**.

# The Girvan-Newman algorithm

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

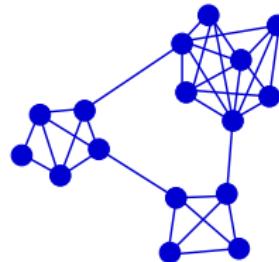
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The inter-community links have high **betweenness**:



## The algorithm

- 1) calculate the betweenness for the links,
- 2) delete the one with the highest  $b$ ,  
if disconnected parts emerge, update the dendrogram,  
→ go back to step 1)

# The Girvan-Newman algorithm

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile  
profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

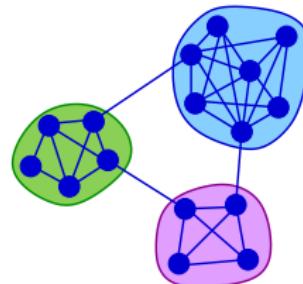
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The inter-community links have high **betweenness**:



## The algorithm

- 1) calculate the betweenness for the links,
- 2) delete the one with the highest  $b$ ,  
if disconnected parts emerge, update the dendrogram,  
→ go back to step 1)

# The Girvan-Newman algorithm

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

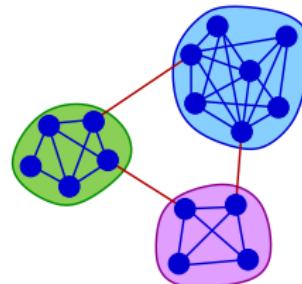
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The inter-community links have high **betweenness**:



## The algorithm

- 1) calculate the betweenness for the links,
- 2) delete the one with the highest  $b$ ,  
if disconnected parts emerge, update the dendrogram,  
→ go back to step 1)

# The Girvan-Newman algorithm

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

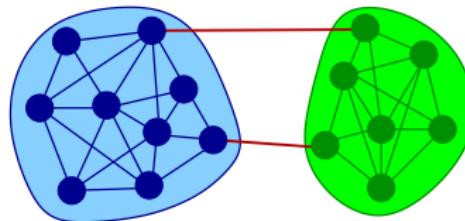
Overlapping communities

The Clique

Percolation Method

The Infomap method

The recalculations of the betweenness after each cut is crucial:



We may have more than one bridging link between two communities, and if one of these bridges is used by most of the shortest paths between the two modules, the betweenness of the secondary bridges will not be outstandingly high. However, if we remove the "main" bridge first, the betweenness of the secondary bridge will suddenly increase, and then we can find it just as we had found the "main" bridge in the previous step.

# Modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

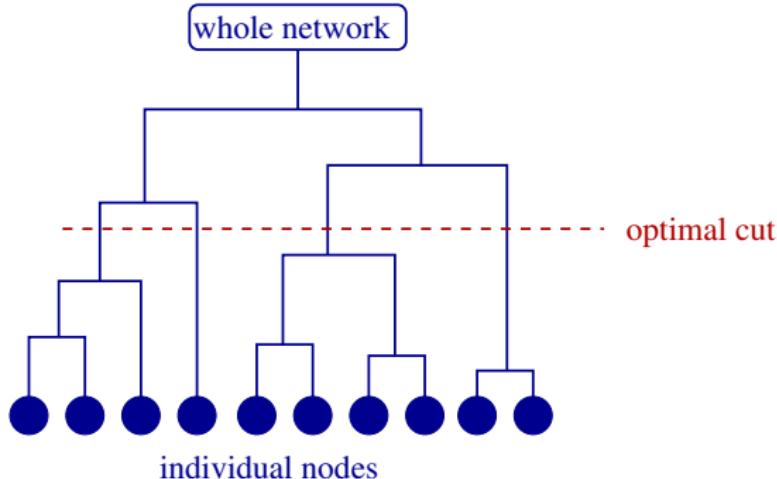
Overlapping communities

The Clique

Percolation Method

The Infomap method

- Where is the optimal cut of the dendrogram?



- In order to find the optimal cut, Girvan and Newman introduced a quantity called modularity, which is measuring the quality of the partitioning of the system communities. Once we have defined such a quantity, the optimal cut is found where we have a maximum in the modularity.

# Modularity

- How to measure the quality of individual communities?

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

# Modularity

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

## Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

## Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?  
→ Let's compare them to random partitions!

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?
  - Let's compare them to random partitions!
- “High quality” communities: much more links than expected at random.
- “Low quality” communities: as many links (or even less than) as expected at random.

# Modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

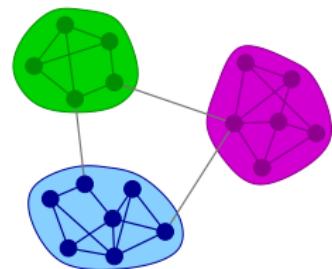
Overlapping communities

The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?
  - Let's compare them to random partitions!
    - “High quality” communities: much more links then expected at random.
    - “Low quality” communities: as many links (or even less then) as expected at random.
- Technically it is simpler to keep the community memberships fixed, and rewire the network at random.



# Modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

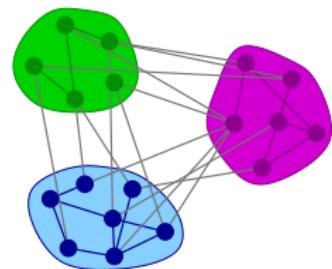
Overlapping communities

The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?
  - Let's compare them to random partitions!
    - “High quality” communities: much more links then expected at random.
    - “Low quality” communities: as many links (or even less then) as expected at random.
- Technically it is simpler to keep the community memberships fixed, and rewire the network at random.



# Modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

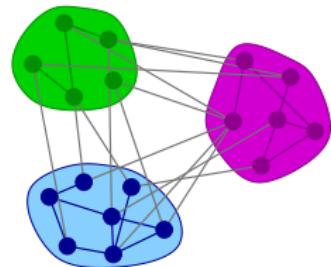
The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?
- Let's compare them to random partitions!
  - “High quality” communities: much more links then expected at random.
  - “Low quality” communities: as many links (or even less then) as expected at random.

- Technically it is simpler to keep the community memberships fixed, and rewire the network at random.
- What should be the reference random graph model?



# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

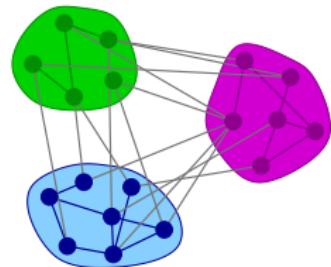
The Clique

Percolation Method

The Infomap method

- How to measure the quality of individual communities?
  - Let's compare them to random partitions!
  - “High quality” communities: much more links then expected at random.
  - “Low quality” communities: as many links (or even less then) as expected at random.

- Technically it is simpler to keep the community memberships fixed, and rewire the network at random.
- What should be the reference random graph model?
  - The configuration model.



# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

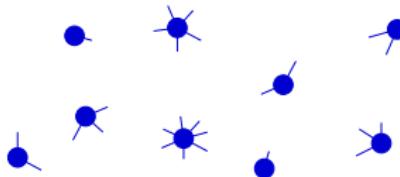
Overlapping communities

The Clique

Percolation Method

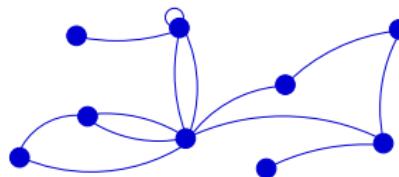
The Infomap method

- The configuration model:



# Modularity

- The configuration model:



Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

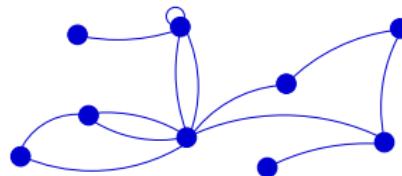
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

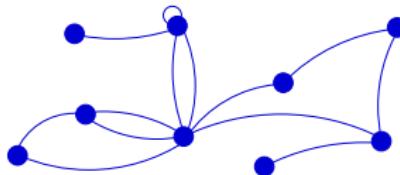
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$

# Modularity

## Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

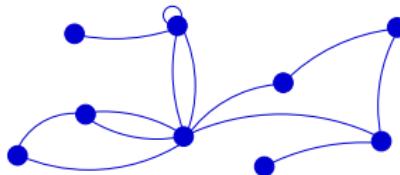
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?  
$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$
- Expected number of links inside community  $\alpha$ ?

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

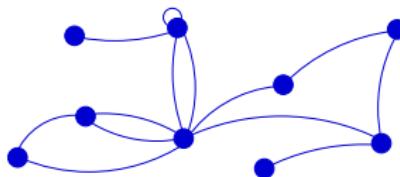
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$

- Expected number of links inside community  $\alpha$ ?

- the expected number of links from between  $i$  and other community members:

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

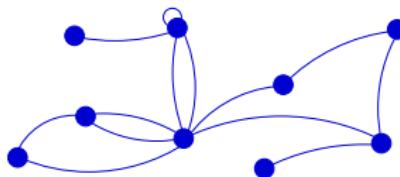
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$

- Expected number of links inside community  $\alpha$ ?

- the expected number of links from between  $i$  and other community members:  $M_{i\alpha} = k_i \sum_{j \in \alpha} \frac{k_j}{2L}$ .

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

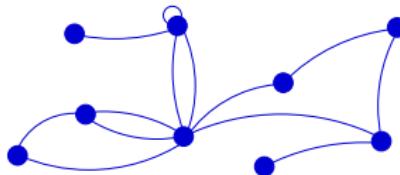
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$

- Expected number of links inside community  $\alpha$ ?

- the expected number of links from between  $i$  and other community members:  $M_{i\alpha} = k_i \sum_{j \in \alpha} \frac{k_j}{2L}.$

- summing over all community members gives twice the number of links in  $\alpha$ :

$$M_\alpha = \frac{1}{2} \sum_{i \in \alpha} k_i \sum_{j \in \alpha} \frac{k_j}{2L}.$$

# Modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

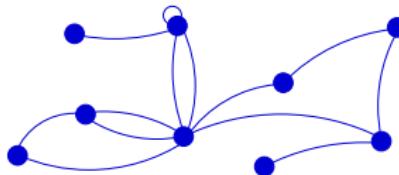
Overlapping communities

The Clique

Percolation Method

The Infomap method

- The configuration model:



- Probability for nodes  $i$  and  $j$  to be connected?

$$\mathcal{P}(i - j) = \frac{k_i k_j}{2L}$$

- Expected number of links inside community  $\alpha$ ?

- the expected number of links from between  $i$  and other community members:  $M_{i\alpha} = k_i \sum_{j \in \alpha} \frac{k_j}{2L}$ .

- summing over all community members gives twice the number of links in  $\alpha$ :

$$M_\alpha = \frac{1}{2} \sum_{i \in \alpha} k_i \sum_{j \in \alpha} \frac{k_j}{2L}.$$

- Expected fraction of links in  $\alpha$ :

$$\frac{M_\alpha}{M} = \frac{1}{2L} \sum_{i \in \alpha} k_i \sum_{j \in \alpha} \frac{k_j}{2L} = \sum_{i \in \alpha} \frac{k_i}{2L} \sum_{j \in \alpha} \frac{k_j}{2L} = \left[ \sum_{i \in \alpha} \frac{k_i}{2L} \right]^2$$

# Modularity

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- The modularity of community  $\alpha$ :

$$Q_\alpha \equiv \underbrace{\frac{l_\alpha}{L}}_{\text{observed}} - \underbrace{\left(\frac{k_\alpha}{2L}\right)^2}_{\text{expected}}$$

$l_\alpha$  : num. of links in  $\alpha$ ,  
 $k_\alpha = \sum_{i \in \alpha} k_i$

- The modularity of the whole community partition:

$$Q = \sum_\alpha Q_\alpha = \sum_\alpha \left[ \frac{l_\alpha}{L} - \left( \frac{k_\alpha}{2L} \right)^2 \right]$$

- A widely used, equivalent formulation of  $Q$ :

$$Q = \frac{1}{2L} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2L} \right) \delta(\alpha_i, \alpha_j)$$

# Modularity

## Illustration

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

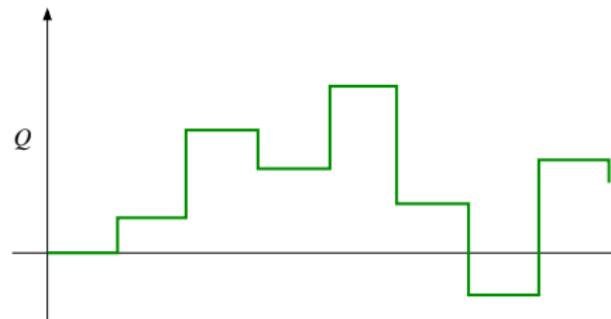
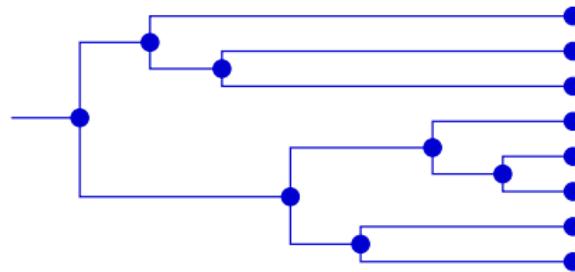
#### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

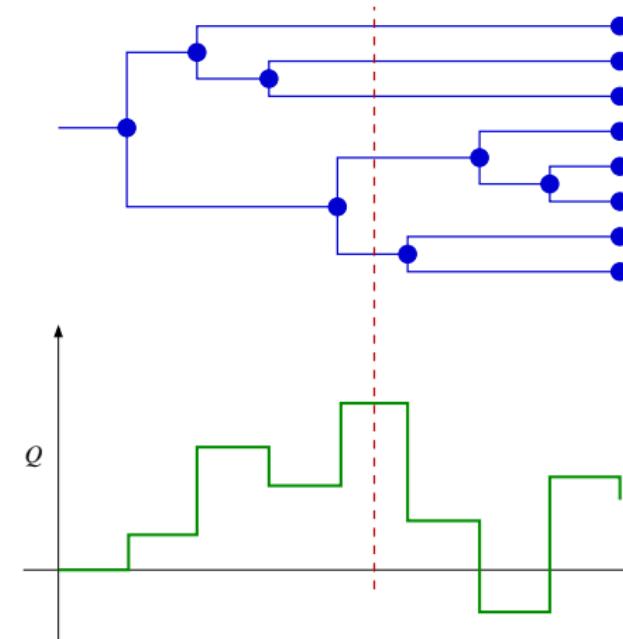


## Modularity

## Illustration

## Motifs and communities

## Modularity



# Modularity

## Illustration

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

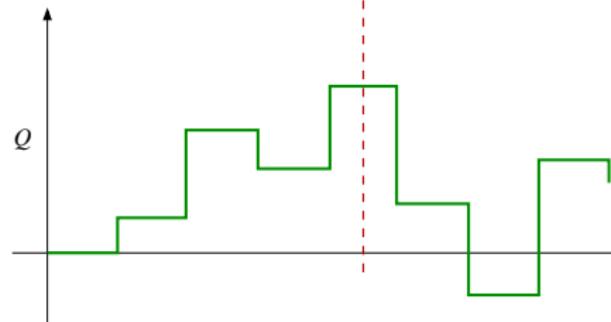
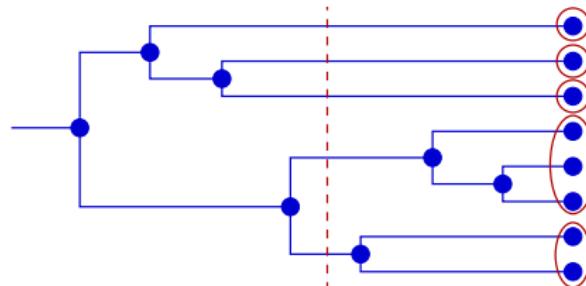
#### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# Modularity

## Illustration

### Motifs and communities

#### Motifs

##### Definition

##### z-score

##### Motif significance profile

#### Communities

##### Examples

##### Hierarchical clustering

##### The Girvan-Newman algorithm

#### Modularity

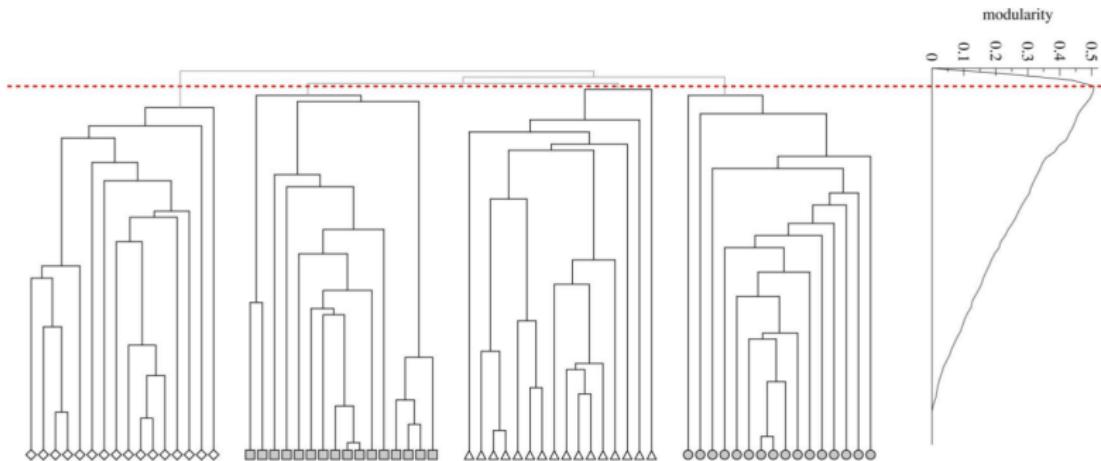
##### Overlapping communities

##### The Clique

##### Percolation Method

##### The Infomap method

Results on a synthetic test network consisting of 4 denser communities that are connected to each other in a sparser manner:



# Maximal Q partitions found by GN algorithm

## Illustration

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

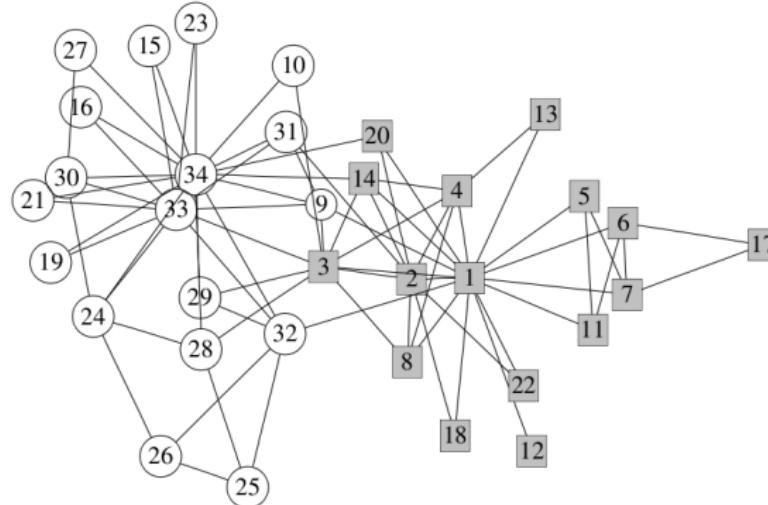
Overlapping communities

The Clique

Percolation Method

The Infomap method

Results on the famous Zakary's karate club, which split into two communities. The colours and the shapes indicate the results from the Girvan-Newman algorithm and the "ground truth" communities based on sociologists studying the club.



# Modularity optimization

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- If the modularity is used to determine the quality of the partitioning, why not optimize  $Q$  directly?
  - Finding the partition with maximal  $Q$  is an NP-complete problem.
  - Still, efficient approximate algorithms can be designed.

# Modularity optimization

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- If the modularity is used to determine the quality of the partitioning, why not optimize  $Q$  directly?
- Finding the partition with maximal  $Q$  is an NP-complete problem.
  - Still, efficient approximate algorithms can be designed.

# Modularity optimization

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- If the modularity is used to determine the quality of the partitioning, why not optimize  $Q$  directly?
  - Finding the partition with maximal  $Q$  is an NP-complete problem.
  - Still, efficient approximate algorithms can be designed.

# Modularity optimization

## Greedy approach

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

### Greedy algorithm

- Initially every node is in a separate community.
- At each step join a pair of communities, choose according to the maximal increase in  $Q$ .
- If  $Q$  cannot be increased further, stop.

M. E. J. Newman, *Phys. Rev. E* **69**, 066133, (2004)

# Modularity optimization

## Simulated annealing

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

### Simulated annealing for Q

- Start from a random partition.
- Either try moving nodes between communities, or try merging/splitting communities at random.
- The acceptance probability for a given move:

$$p = \begin{cases} 1 & \text{if } \Delta Q \geq 0 \\ \exp\left(\frac{\Delta Q}{T}\right) & \text{if } \Delta Q < 0 \end{cases}$$

- Slowly reduce  $T$  and the system settles in a local optimum.

Guimerà et al., *Phys. Rev. E* **70**, 025101, (2004)

C. P. Massen and J. P. Doyle, *Phys. Rev. E* **71**, 046101, (2005)

# Modularity optimization

## Other approaches

### Motifs and communities

#### Motifs

Definition

$z$ -score

Motif significance profile

#### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- Spectral methods,  
M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103**, 8577 (2006)
- Extremal optimization,  
J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005)
- etc.

# Drawbacks of modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- **Resolution limit:** communities with  $I_\alpha < \sqrt{L/2}$  cannot be resolved.

S. Fortunato and M. Barthélemy, *PNAS* **104**, 36 (2007)

→ Solution:

- local methods,
- multi-resolution methods.

- Overlaps between the communities are not allowed.

# Drawbacks of modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- **Resolution limit:** communities with  $I_\alpha < \sqrt{L/2}$  cannot be resolved.

S. Fortunato and M. Barthélemy, *PNAS* **104**, 36 (2007)

→ Solution:

- local methods,
- multi-resolution methods.

- Overlaps between the communities are not allowed.

# Drawbacks of modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- **Resolution limit:** communities with  $l_\alpha < \sqrt{L/2}$  cannot be resolved.

S. Fortunato and M. Barthélemy, *PNAS* **104**, 36 (2007)

→ Solution:

- local methods,
- multi-resolution methods.

- Overlaps between the communities are not allowed.

# Drawbacks of modularity

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

### Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- **Resolution limit:** communities with  $I_\alpha < \sqrt{L/2}$  cannot be resolved.

S. Fortunato and M. Barthélemy, *PNAS* **104**, 36 (2007)

→ Solution:

- local methods,
- multi-resolution methods.

- Overlaps between the communities are not allowed.

# Drawbacks of modularity

## Motifs and communities

### Motifs

Definition

z-score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- **Resolution limit:** communities with  $I_\alpha < \sqrt{L/2}$  cannot be resolved.

S. Fortunato and M. Barthélemy, *PNAS* **104**, 36 (2007)

→ Solution:

- local methods,
- multi-resolution methods.

- Overlaps between the communities are not allowed.

# Overlaps between communities

## Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

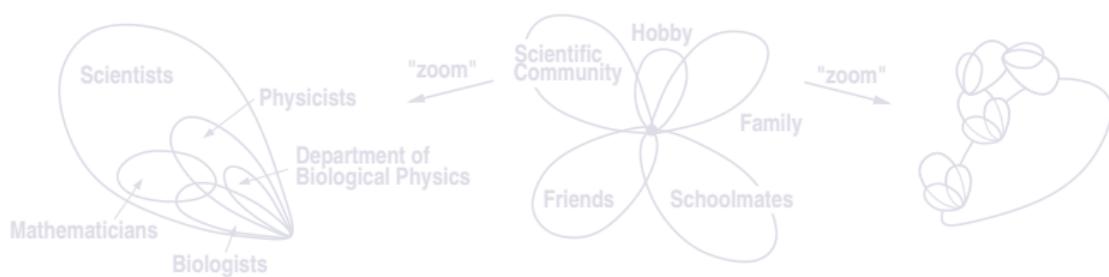
Percolation Method

The Infomap method

Communities can overlap:

- People can be members of different social groups at the same time
- A protein can be part of more than one functional unit

Furthermore, they can be nested as well:



# Overlaps between communities

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

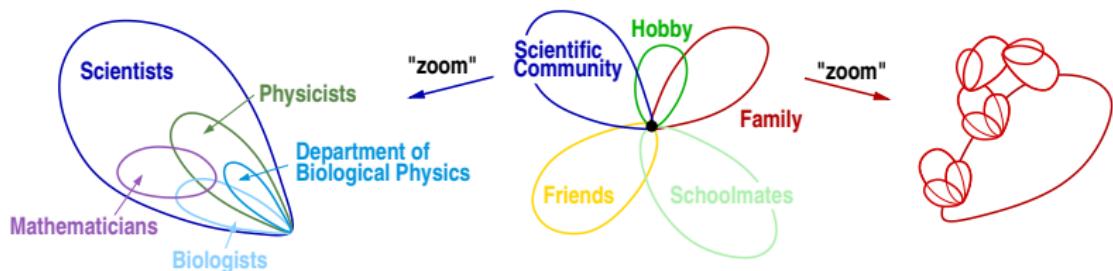
Percolation Method

The Infomap method

Communities can overlap:

- People can be members of different social groups at the same time
- A protein can be part of more than one functional unit

Furthermore, they can be nested as well:



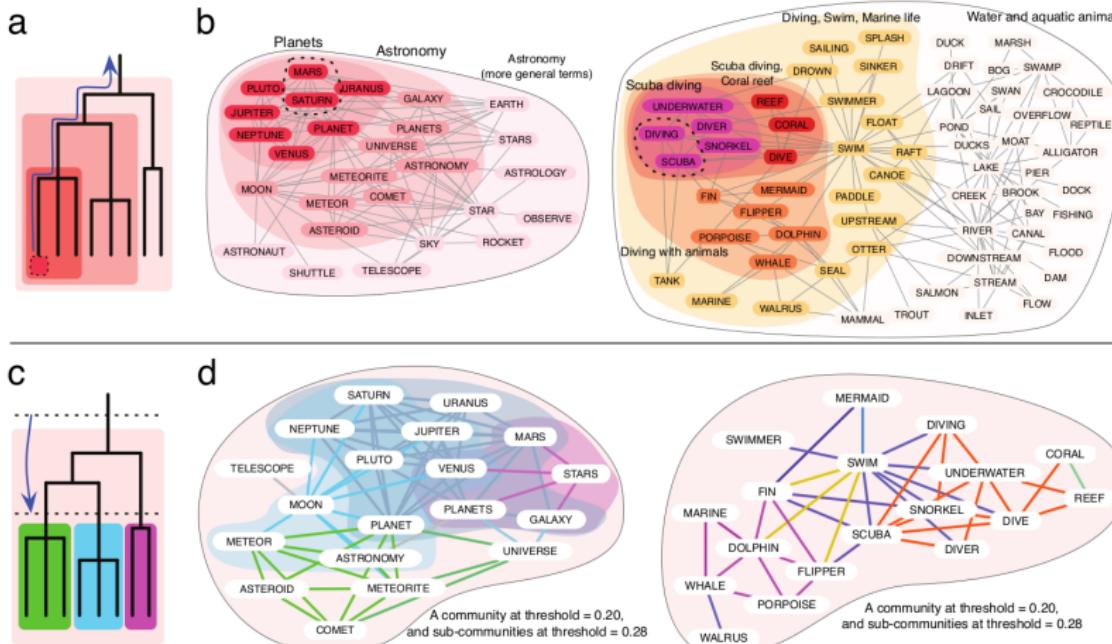
## Nested communities

## An example

## Motifs and communities

## Modularity Overlapping

### The Clique



Y.-Y. Ahn, J. P. Bagrow and S. Lehmann, *Nature* **466**, 761 (2010)

# The Clique Percolation Method (CPM)

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique Percolation Method

The Infomap method

The CPM is a **local** method allowing **overlaps** between the communities.

## Definitions

- **$k$ -clique**: a complete (fully connected) subgraph of  $k$  nodes.
- **$k$ -clique adjacency**: two  $k$ -cliques are adjacent if they share  $k - 1$  nodes, *i.e.*, if they differ only in a single node.
- **$k$ -clique community**: the union of  $k$ -cliques that can be reached from one to the other through a sequence of adjacent  $k$ -cliques.



G. Palla, I. Derényi, I. Farkas and T. Vicsek, *Nature* **435**, 814 (2005)

# The Clique Percolation Method (CPM)

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique Percolation Method

The Infomap method

The CPM is a **local** method allowing **overlaps** between the communities.

## Definitions

- **$k$ -clique**: a complete (fully connected) subgraph of  $k$  nodes.
- **$k$ -clique adjacency**: two  $k$ -cliques are adjacent if they share  $k - 1$  nodes, i.e., if they differ only in a single node.
- **$k$ -clique community**: the union of  $k$ -cliques that can be reached from one to the other through a sequence of adjacent  $k$ -cliques.



G. Palla, I. Derényi, I. Farkas and T. Vicsek, *Nature* **435**, 814 (2005)

# The Clique Percolation Method (CPM)

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

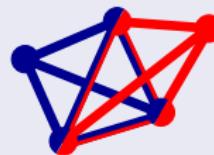
The Clique Percolation Method

The Infomap method

The CPM is a **local** method allowing **overlaps** between the communities.

## Definitions

- **$k$ -clique**: a complete (fully connected) subgraph of  $k$  nodes.
- **$k$ -clique adjacency**: two  $k$ -cliques are adjacent if they share  $k - 1$  nodes, i.e., if they differ only in a single node.
- **$k$ -clique community**: the union of  $k$ -cliques that can be reached from one to the other through a sequence of adjacent  $k$ -cliques.



G. Palla, I. Derényi, I. Farkas and T. Vicsek, *Nature* **435**, 814 (2005)

# The Clique Percolation Method (CPM)

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

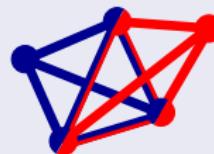
The Clique Percolation Method

The Infomap method

The CPM is a **local** method allowing **overlaps** between the communities.

## Definitions

- **$k$ -clique**: a complete (fully connected) subgraph of  $k$  nodes.
- **$k$ -clique adjacency**: two  $k$ -cliques are adjacent if they share  $k - 1$  nodes, *i.e.*, if they differ only in a single node.
- **$k$ -clique community**: the union of  $k$ -cliques that can be reached from one to the other through a sequence of adjacent  $k$ -cliques.



G. Palla, I. Derényi, I. Farkas and T. Vicsek, *Nature* **435**, 814 (2005)

# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

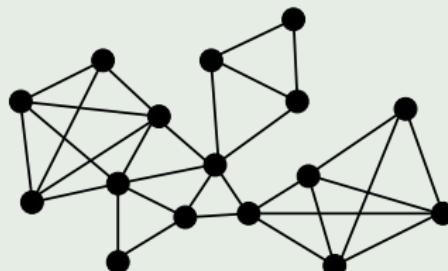
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

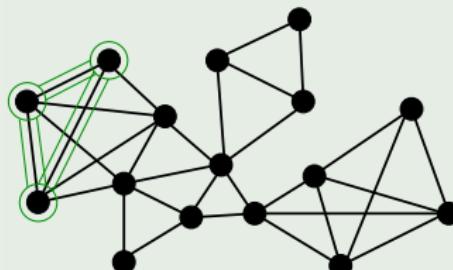
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile  
profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

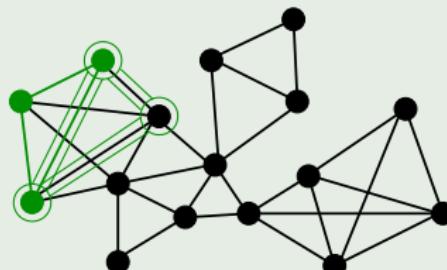
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

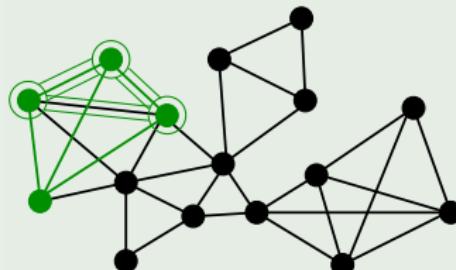
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

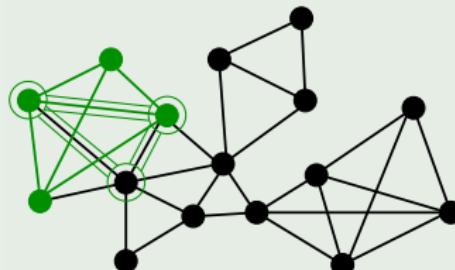
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

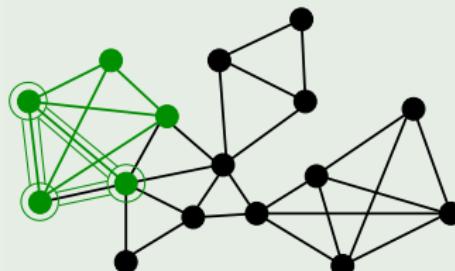
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

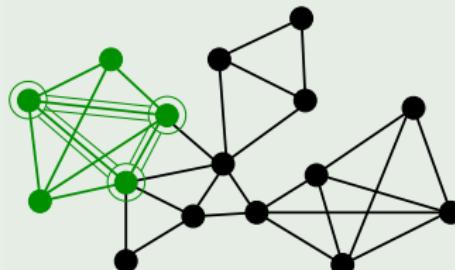
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

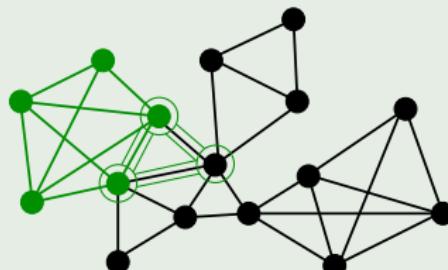
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

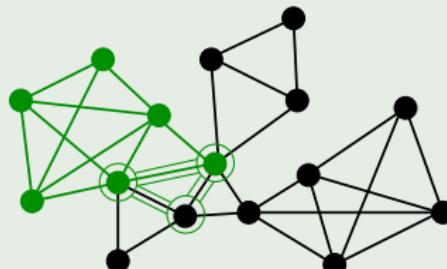
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

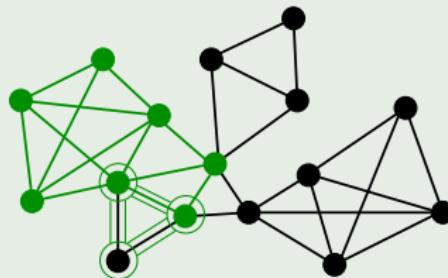
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

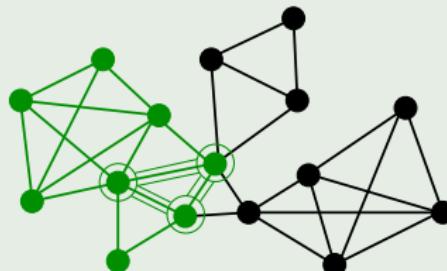
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

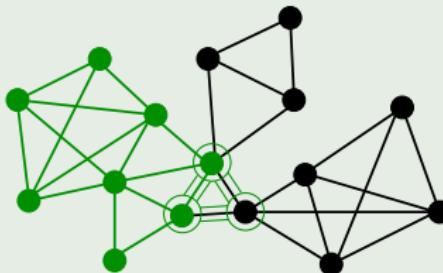
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

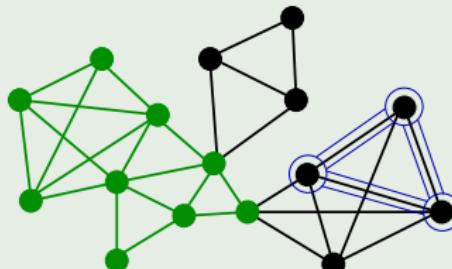
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

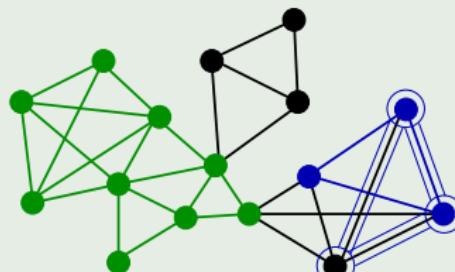
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

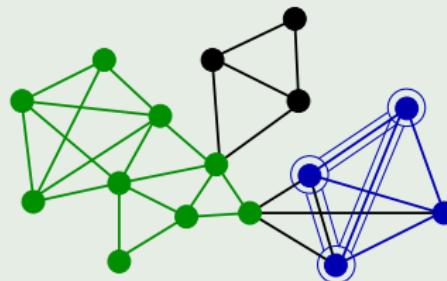
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

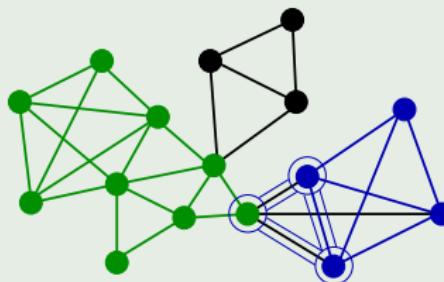
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

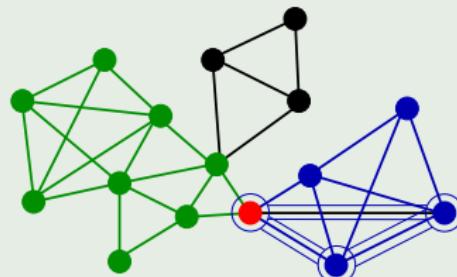
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

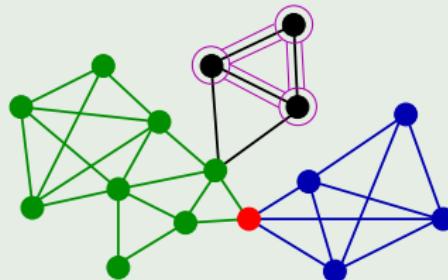
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

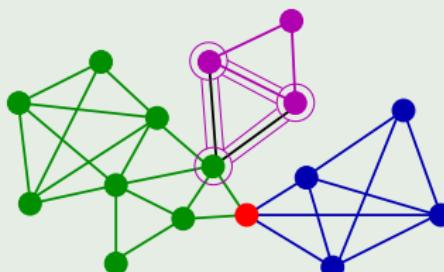
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

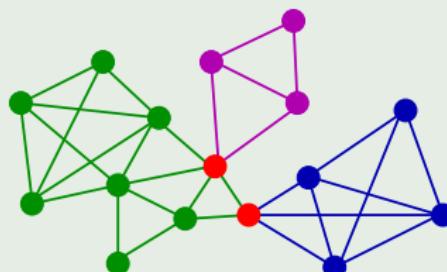
Overlapping communities

The Clique

Percolation Method

The Infomap method

### Illustration:



# CPM

(continued)

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

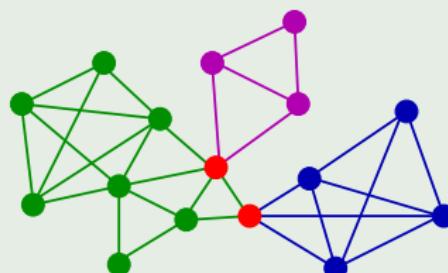
Overlapping communities

The Clique

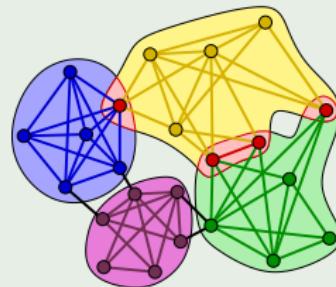
Percolation Method

The Infomap method

Illustration:



same at  $k = 4$ :



# Communities of a given node

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

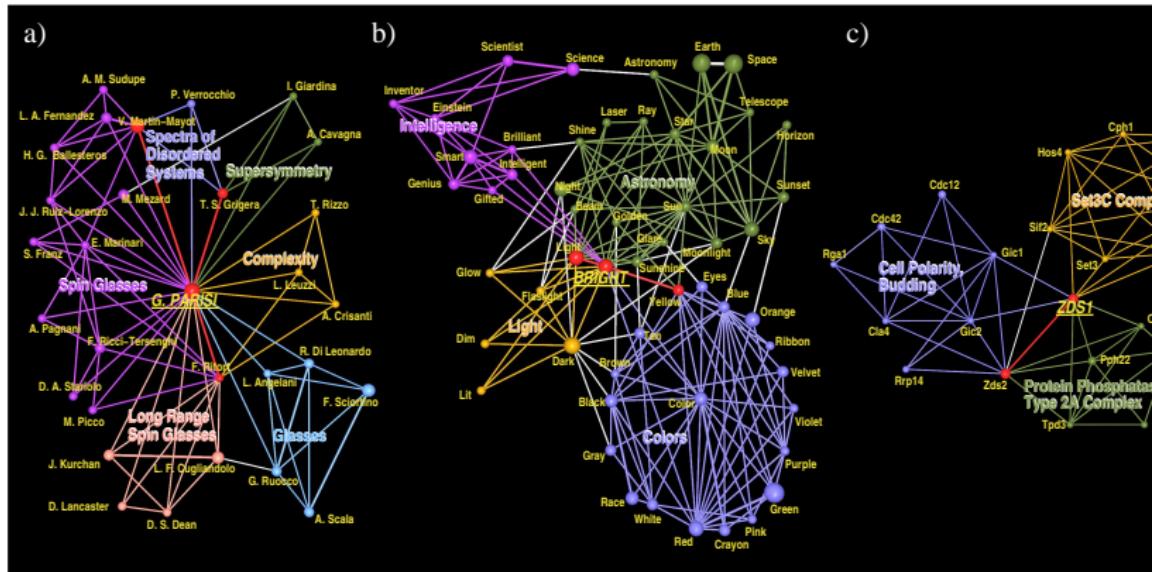
Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method



# Communities of a given node

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

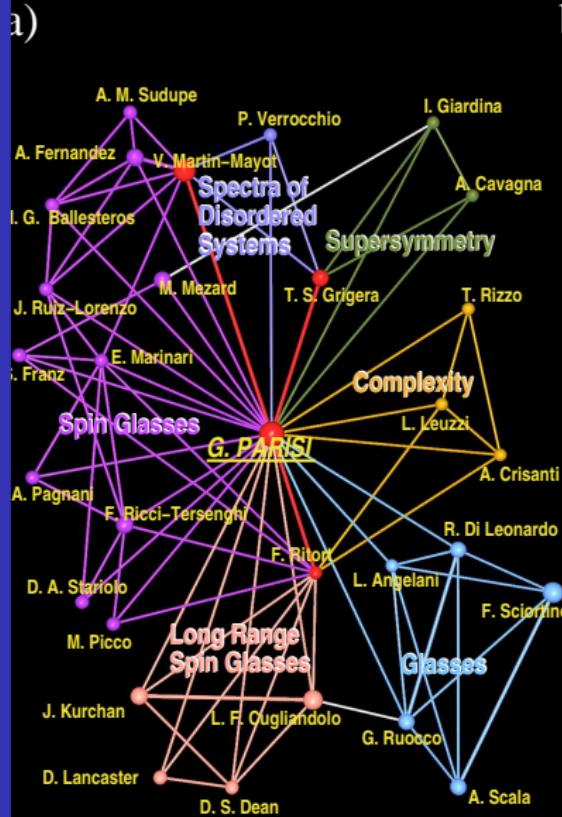
Overlapping communities

The Clique

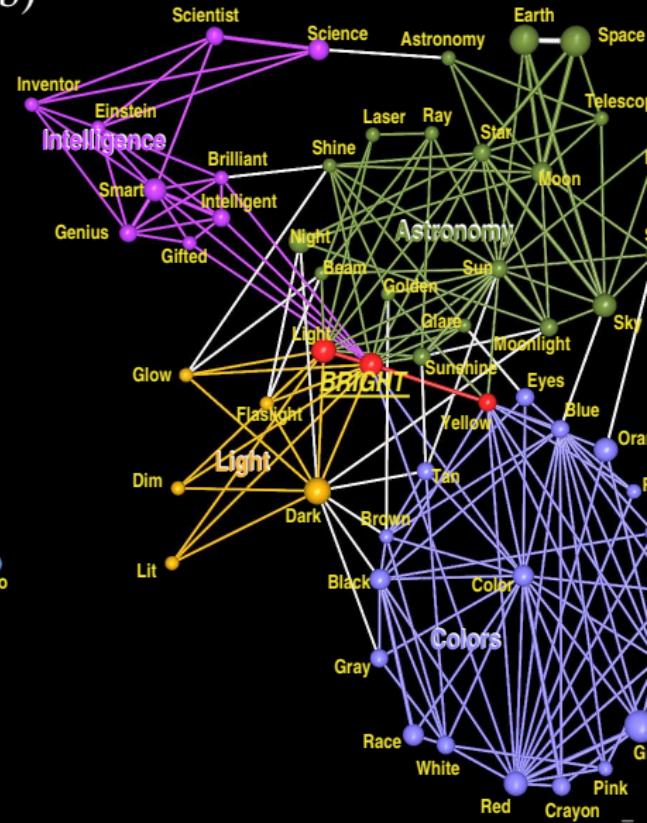
Percolation Method

The Infomap method

a)



b)



# Percolation transition

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## What about the community structure of a whole network?

- High  $k$  and  $w^*$ : small, dispersed communities.
- Low  $k$  and  $w^*$ : a giant community, smearing out the details by swallowing many smaller communities.

### Optimal $k$ -clique size and $w^*$

Where the community structure is as highly structured as possible: just above the critical point of the appearance of a giant  $k$ -clique community.

# Percolation transition

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

## What about the community structure of a whole network?

- High  $k$  and  $w^*$ : small, dispersed communities.
- Low  $k$  and  $w^*$ : a giant community, smearing out the details by swallowing many smaller communities.

### Optimal $k$ -clique size and $w^*$

Where the community structure is as highly structured as possible: just above the critical point of the appearance of a giant  $k$ -clique community.

# Percolation transition

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

What about the community structure of a whole network?

- High  $k$  and  $w^*$ : small, dispersed communities.
- Low  $k$  and  $w^*$ : a giant community, smearing out the details by swallowing many smaller communities.

Optimal  $k$ -clique size and  $w^*$

Where the community structure is as highly structured as possible: just above the critical point of the appearance of a giant  $k$ -clique community.

# Percolation transition

Motifs and communities

Motifs

Definition

$z$ -score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

What about the community structure of a whole network?

- High  $k$  and  $w^*$ : small, dispersed communities.
- Low  $k$  and  $w^*$ : a giant community, smearing out the details by swallowing many smaller communities.

Optimal  $k$ -clique size and  $w^*$

Where the community structure is as highly structured as possible: just above the critical point of the appearance of a giant  $k$ -clique community.

# CFinder

Motifs  
and  
communities

Motifs

Definition

z-score

Motif significance  
profile

Communities

Examples

Hierarchical  
clustering

The Girvan-Newman  
algorithm

Modularity

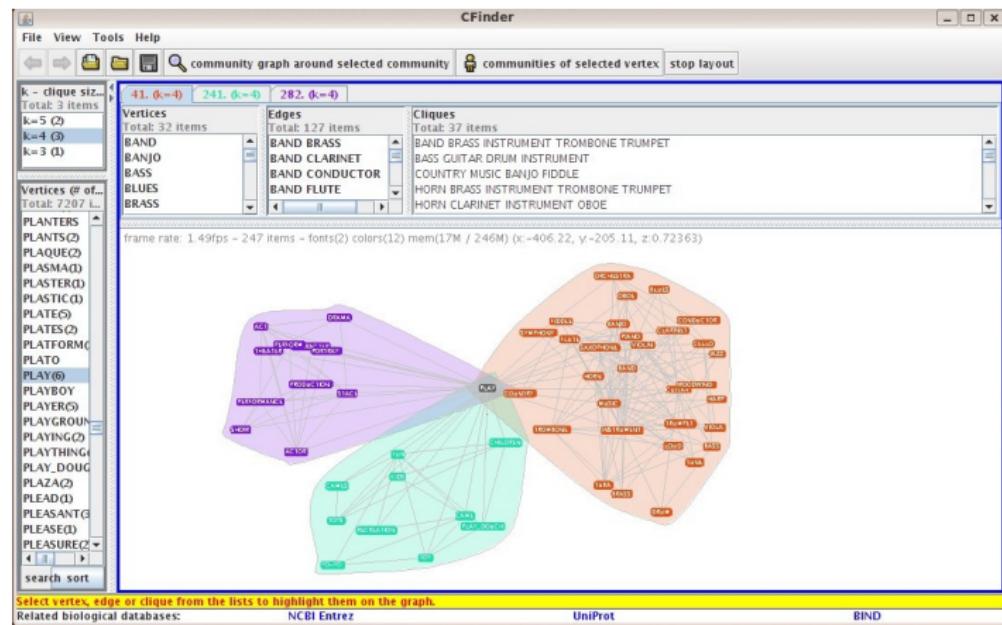
Overlapping  
communities

The Clique

Percolation Method

The Infomap method

A program package for community finding and visualising based on the  
CPM: <http://www.cfinder.org>



# Infomap method

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

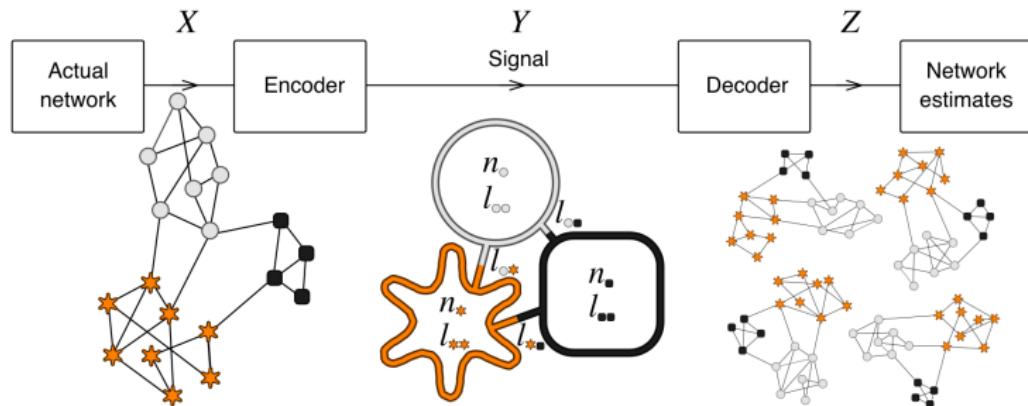
Overlapping communities

The Clique

Percolation Method

The Infomap method

Another approach providing high quality communities is the Infomap method.  
Basic idea:



M. Rosvall and C. T. Bergstrom, *PNAS* **104**, 7327 (2007)

# Infomap method

Motifs and communities

Motifs

Definition

z-score

Motif significance profile  
profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

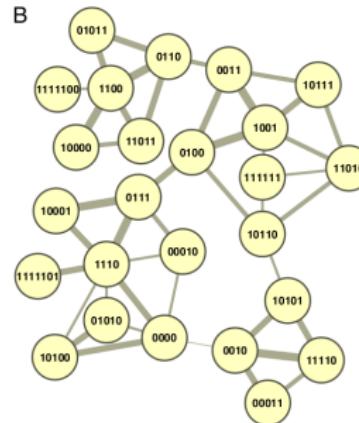
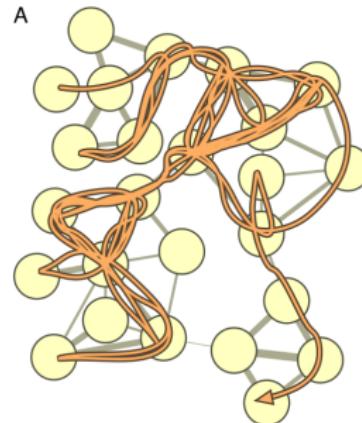
Overlapping communities

The Clique

Percolation Method

The Infomap method

Alternatively:



```
1111100 1100 0110 11011 10000 11011 0110 0011 10111 1001 0011  
1001 1100 0111 1110 0111 10001 0111 1110 0000 1110 1001 0111  
0111 1110 0111 1110 111101 1110 0000 10100 0000 1110 1001 0111  
0100 10110 11010 10111 1001 0100 1001 10111 1001 0100 1001 0100  
0011 0100 0011 0110 1101 0110 0011 0100 1001 10111 0011 0010  
0111 10001 1110 10001 0111 0100 10110 111111 10110 10101 11110  
00011
```

M. Rosvall and C. T. Bergstrom, *PNAS* **105**, 1118 (2008)

# Infomap method

Motifs and communities

Motifs

Definition

z-score

Motif significance profile  
profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

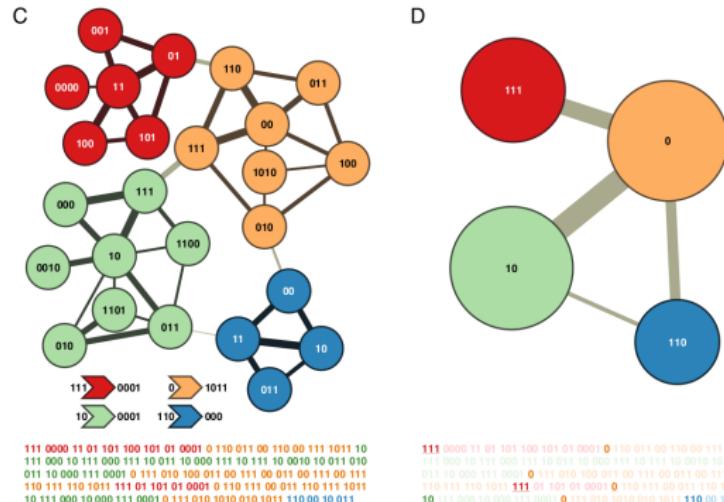
Overlapping communities

The Clique

Percolation Method

The Infomap method

Alternatively:



M. Rosvall and C. T. Bergstrom, *PNAS* **105**, 1118 (2008)

# Infomap method

## Motifs and communities

### Motifs

Definition

$z$ -score

Motif significance profile

### Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

More info, downloadable code and on-line clustering:  
<https://www.mapequation.org/>

MapEquation

Explore the mechanics of the map equation


$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^m p_{\circlearrowleft}^i H(\mathcal{P}^i)$$

Apps »



```
from infomap import Infomap
im = Infomap()
im.read_file("ninetriangles.net")
im.add_link(1, 10)
im.run("--two-level --num-trials 5")
print(im.codewordLength)
for node in im.tree:
    if node.is_leaf:
        print(node.node_id, node.module_id)
```

Code »

Publications »

Maps of information flow reveal community structure in complex networks  
Marten Rosvall and Carl T. Bergstrom  
PNAS 105:1118 (2008). ([arXiv:0707.0609](https://arxiv.org/abs/0707.0609))

To comprehend the multipartite organization of large-scale biological and social systems, we introduce a new information-theoretic approach to modular community detection in networks.

# Closing

Motifs and communities

Motifs

Definition

z-score

Motif significance profile

Communities

Examples

Hierarchical clustering

The Girvan-Newman algorithm

Modularity

Overlapping communities

The Clique

Percolation Method

The Infomap method

- Community finding is an important and highly non-trivial problem, with many possible different “solutions”, (i.e., clustering methods).
- Always choose the method which best suits your problem, (i.e., structure and size of the network, pre-conceptions about communities , etc.).

Further reading on communities (not compulsory):

- Network science book by A.-L. Barabási, chapter 9:  
<http://networksciencebook.com/chapter/9>
- S. Fortunato and D. Hric: Community detection in networks: A user guide.  
*Physics Reports* **659**, 1-44 (2016).
- S. Fortunato: Community detection in graphs.  
*Physics Reports* **486**, 75-174 (2010).