

Expected Nodes: a quality function for the detection of link communities

Noé Gaumont, François Queyroi, Clémence Magnien and
Matthieu Latapy

LIP6 - CNRS & UPMC , Université Pierre et Marie Curie – Sorbonne Universités, Paris,
France.

CompleNet 2015

Summary

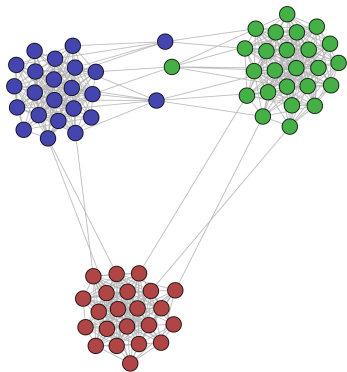
- ① Link community
- ② Expected Nodes : a new quality function
- ③ Tests with LF benchmark
- ④ Conclusion and perspectives

Community Detection

Node community

Link community

Node community



Input :
A graph, $G = (V, E)$.

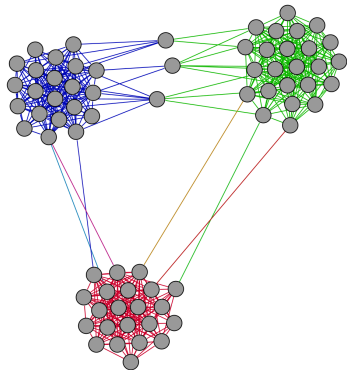
Output :
A partition \mathcal{P} of V .

S. Fortunato.

Community detection in graphs.

Example in a email dataset.
Communities : groups of people.

Link community



Example in a email dataset.
Communities : threads.

Input :

A graph, $G = (V, E)$.

Output :

A partition \mathcal{P} of E .

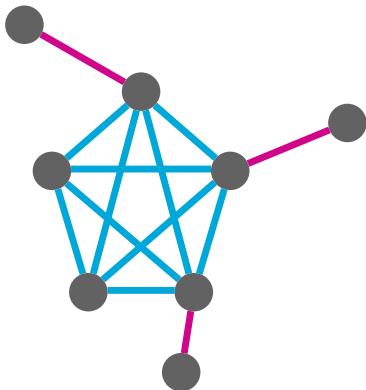
T.S. Evans et R. Lambiotte. **Line graphs, link partitions, and overlapping communities.**

Y.-Y. Ahn, J. P. Bagrow, et S. Lehmann. **Link communities reveal multiscale complexity in networks.**

Expected Nodes : a new quality function

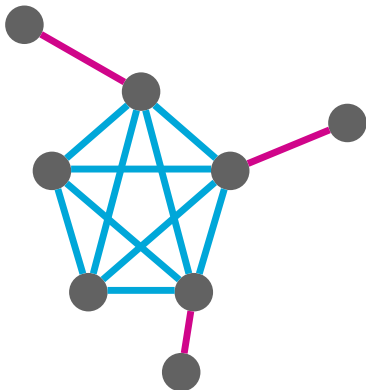
Outline of the quality function

Why is the group of blue links relevant ?



Outline of the quality function

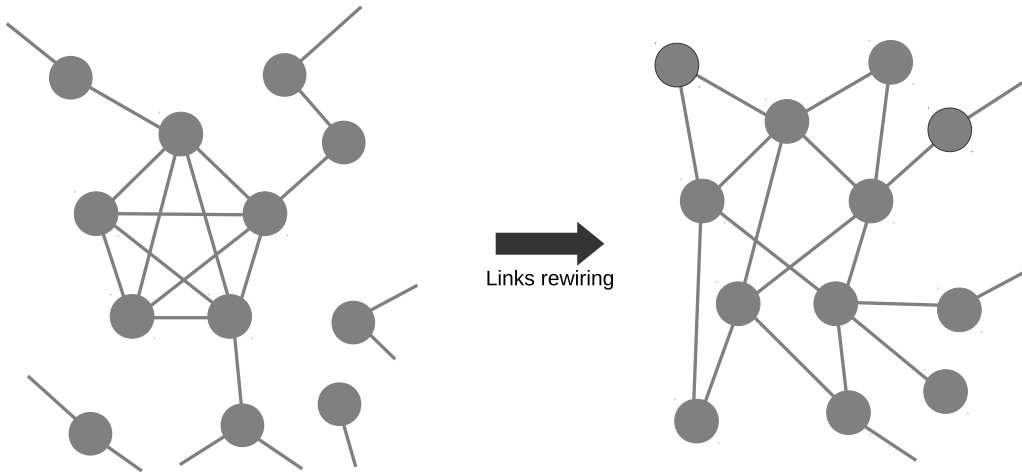
Why is the group of **blue** links relevant ?



Dense **blue links** and sparse **pink links** compare to what could be expected in the **configuration model**.

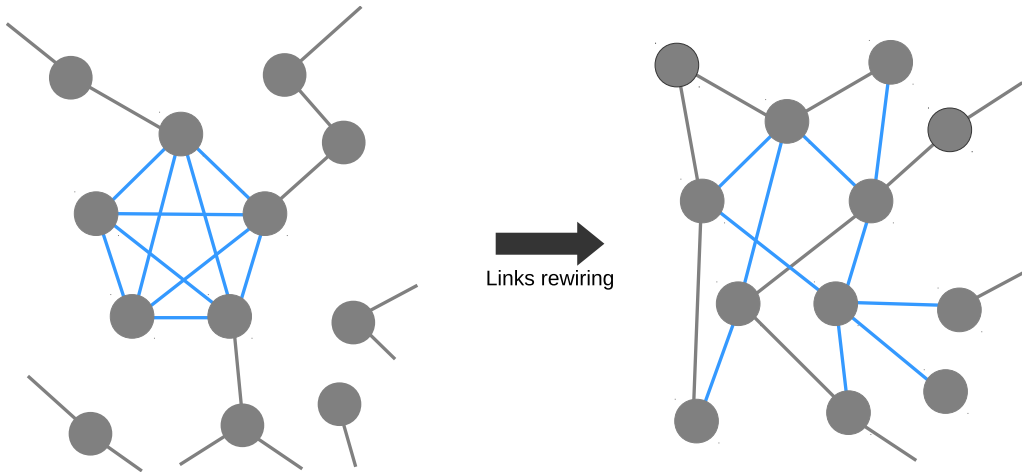
The idea behind Expected Nodes

Compare observed nodes to expected one :



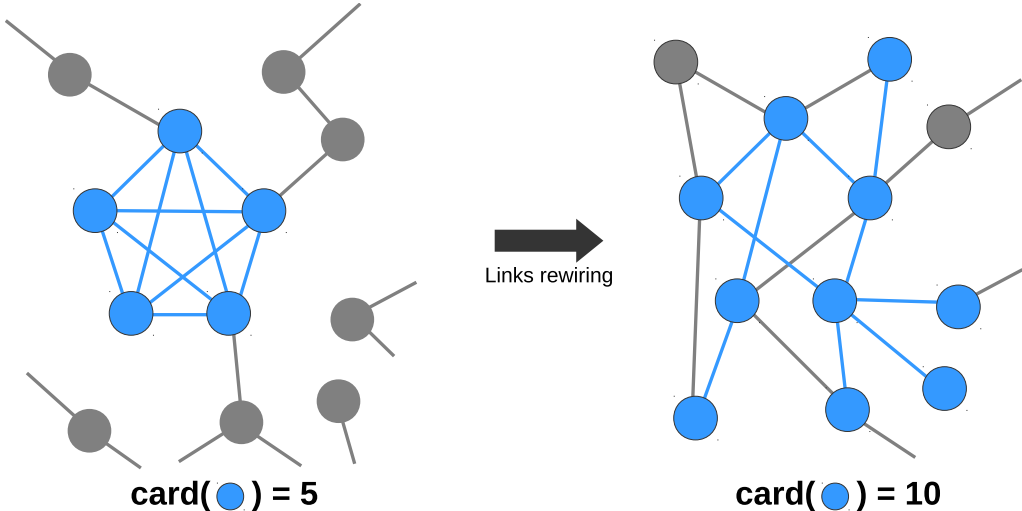
The idea behind Expected Nodes

Compare observed nodes to expected one :



The idea behind Expected Nodes

Compare observed nodes to expected one :



Internal quality function

L : set of **links**, $V(L)$: internal nodes of L .

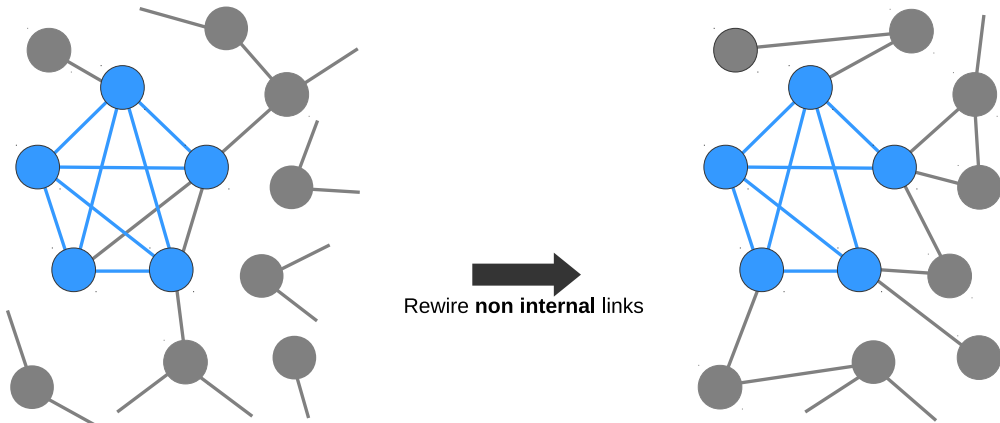
The internal quality of group L is :

$$Q_{in}(L) = \frac{\mathbb{E}[V(L)] - |V(L)|}{\mathbb{E}[V(L)]}$$

$\mathbb{E}[V(L)]$: sum of random variable with hypergeometric distribution.

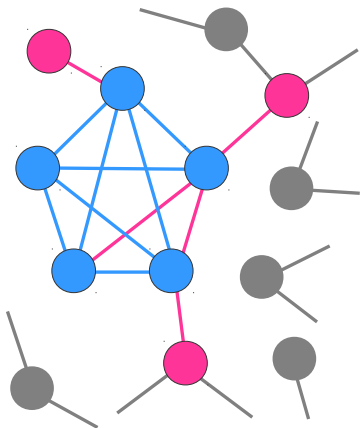
External quality function

Compare adjacent nodes to expected ones :



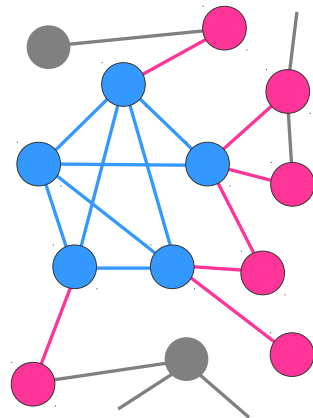
External quality function

Compare adjacent nodes to expected ones :



$$\text{card}(\bullet) = 3$$


 Rewire **non internal** links

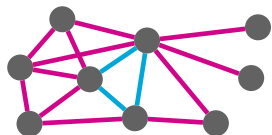


$$\text{card}(\bullet) = 6$$

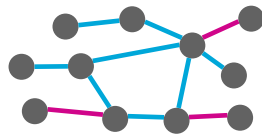
Combining both quality functions

$|L_{out}|$: set of **adjacent links** to L .

$Q_{ext}(L_{out})$ computed in a similar way as $Q_{in}(L)$.



Good internal quality
Bad external quality

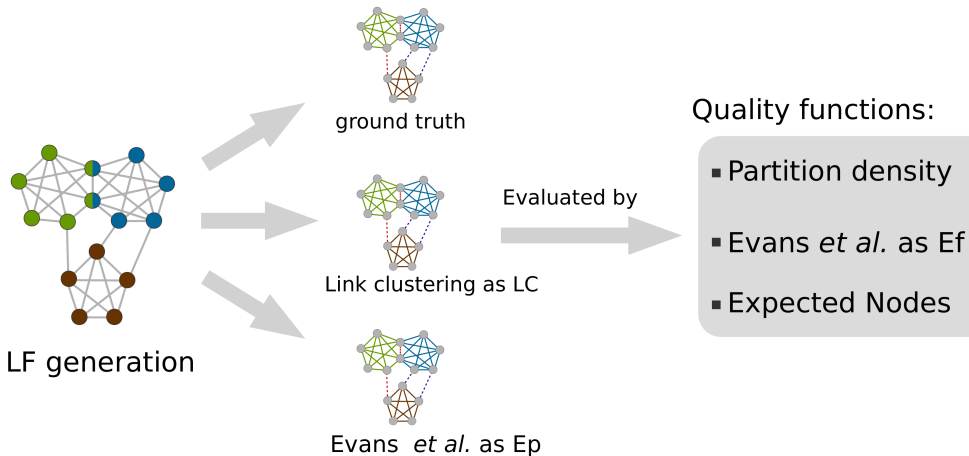


Bad internal quality
Good external quality

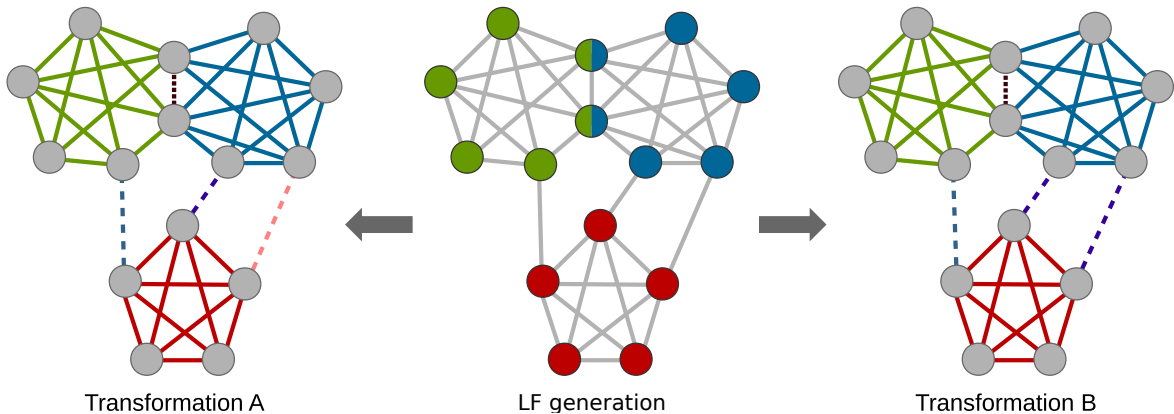
$$Q^*(L) = 2 \frac{|L| Q_{in}(L) + |L_{out}| Q_{ext}(L_{out})}{|L| + |L_{out}|}$$

Tests with LF benchmark

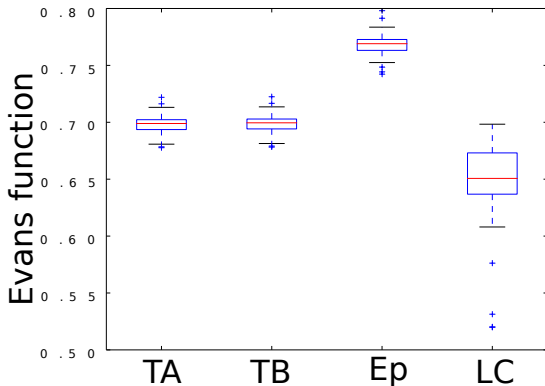
Test method



Ground truth generation



Results for Evans *et al.*



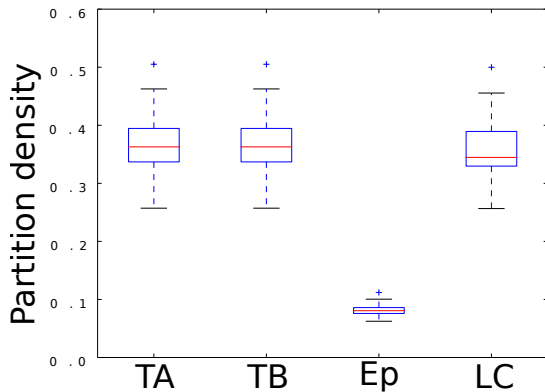
Highlight :

$$Q(TA) < Q(E2)$$

$$Q(TA) = Q(TB)$$

FIGURE – Evaluation of **Ef** from Evans *et al.* on several partitions.

Results for the Partition density

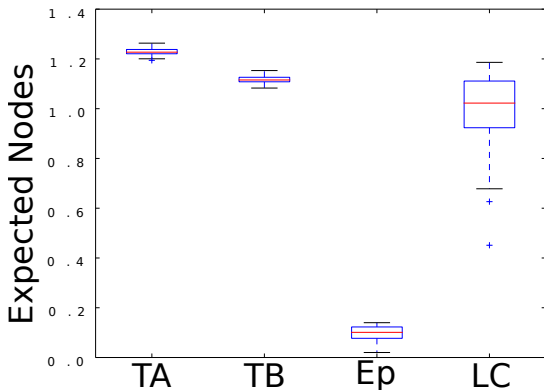


Highlight :

- $Q(TA) \leq Q(LC)$
- $Q(TA) = Q(TB)$

FIGURE – Evaluation of the **partition density** from Ahn *et al.* on several partitions.

Results for Expected Nodes



Highlight :

- $Q(TA) > Q(X)$
- $Q(TA) \neq Q(TB)$

FIGURE – Evaluation of **Expected Nodes** on several partitions.

Conclusion and perspectives

To sum up :

- Consider community of links instead of nodes.
- Definition of *Expected Nodes* to evaluate link partitions.
- On the tests, the ground truth is the best choice only for *Expected Nodes*.

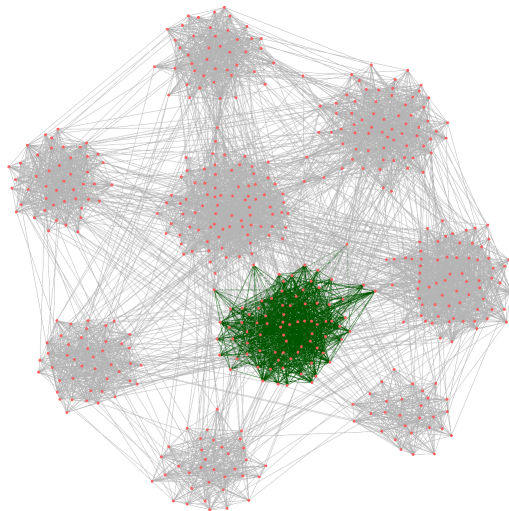
Perspectives

- Design an algorithm for maximizing *Expected Nodes*.
- More detailed comparisons between quality functions

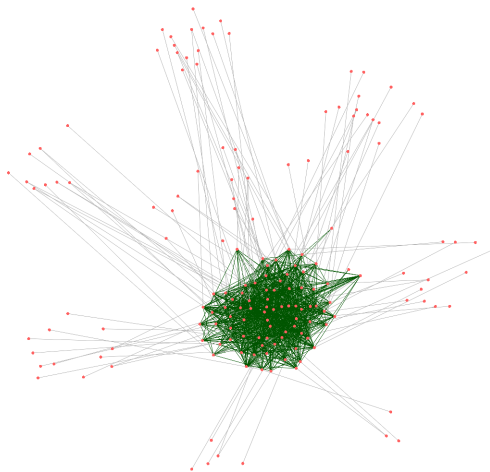
Questions ?

LF generation example

Green group :
a community in the
ground truth



Partition E_p



Partition LC

