



# Relative Evaluation of Partition Algorithms for Complex Networks

**Günce Keziban Orman**


*Galatasaray University C.S. Department / TUBITAK UEKAE*

*keziban.orman@uekae.tubitak.gov.tr*

**Vincent Labatut**

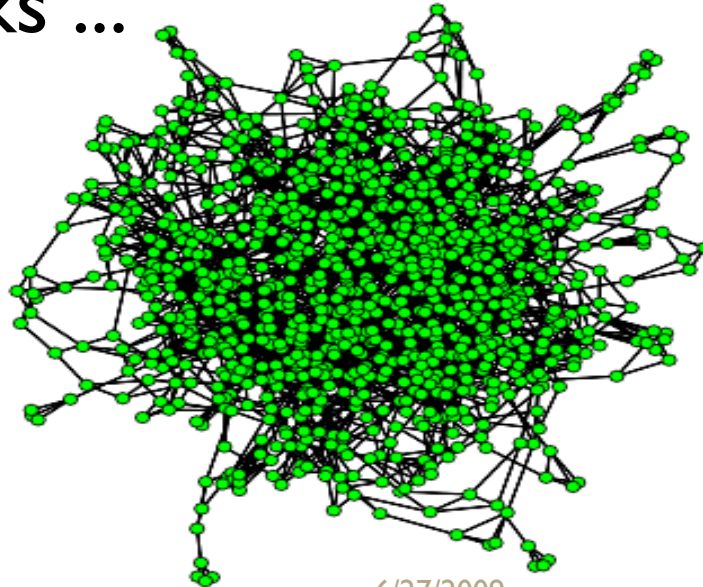
*Galatasaray University C.S Department*

*vlabatut@gsu.edu.tr*

- 
1. Introduction
    1. Complex networks
    2. Community detection
    3. Testing algorithms
  2. Method
    1. Selected algorithms
    2. Lancichinetti *et. al* generative model
    3. Normalized mutual information
  3. Results & Discussions
  4. Conclusion

# I.1 Complex Networks

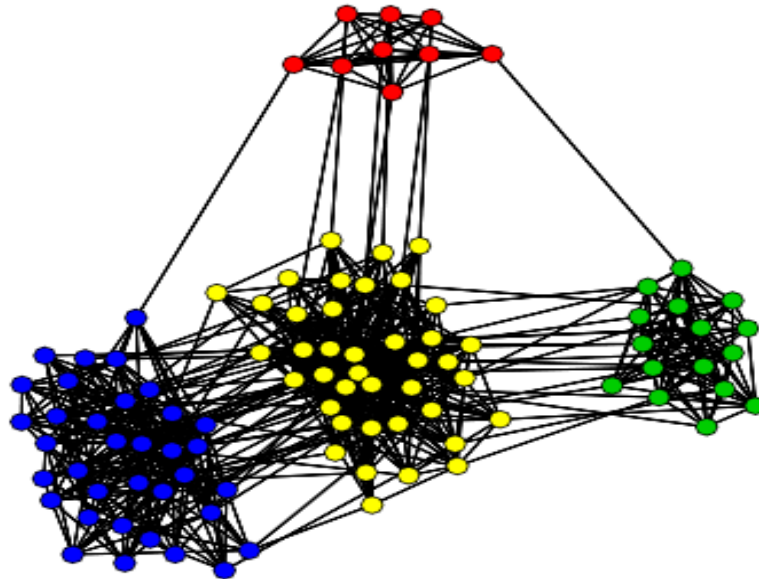
- Large graphs with non-trivial topological features
- Model systems of interacting objects.
- ex: Internet, www, protein web, friendship networks ...



# 1.2.1 Community Detection

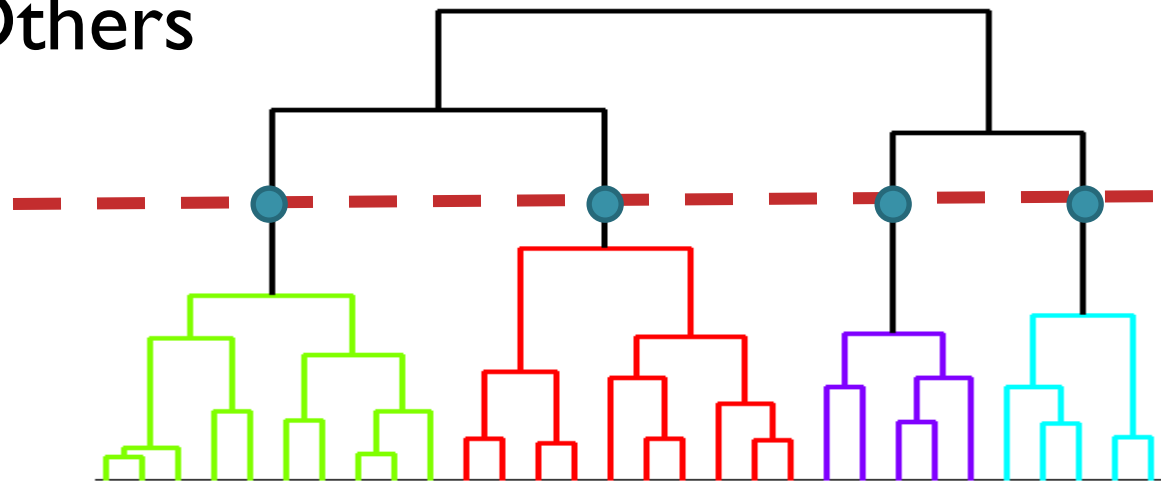
## Definition

- Community: group of nodes with dense inner links and sparse outer links
- Community detection: find the best graph partition according to this definition



# 1.2.2 Community Detection Solution Approaches

- Hierarchical Approaches
  - Divisive vs Agglomerative
- Optimization Approaches
  1. Partition: stochastic vs heuristic
  2. Quality evaluation
- Others



## 1.2.3 Community Detection - Modularity

- Newman's modularity measure:

$$Q = \sum_i (e_{ii} - a_i^2)$$

- $e_{ii}$ : observed fraction of links inside the  $i^{\text{th}}$  community
- $a_i$ : estimation of  $e_{ii}$  under the hypothesis of uniformly randomly distributed links.
- Values:
  - $Q=0$ : networks without community structure and/or random partition
  - $Q \approx 1$ : network with strong community structure and good partition
  - $Q \in [0.3, 0.7]$  is generally considered a good result

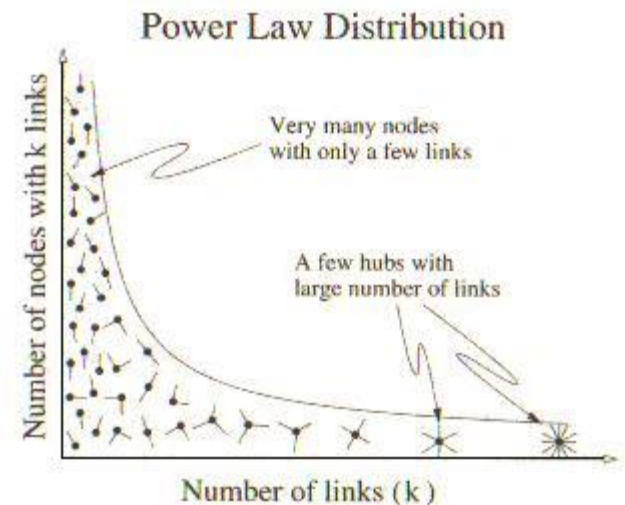
# I.3 Testing Algorithms

	<b>Real networks</b>	<b>Artificial Networks</b>
<b>Size</b>	Usually small	No theoretical limit
<b>Network properties</b>	Uncontrollable, depends on the modeled system	depend on the generation model parameters
<b>Construction</b>	Expensive and/or difficult to build	Computer generated
<b>Community structure</b>	Possibly subjective or unknown community structure	Communities created and controlled

# I.3 Testing Algorithms

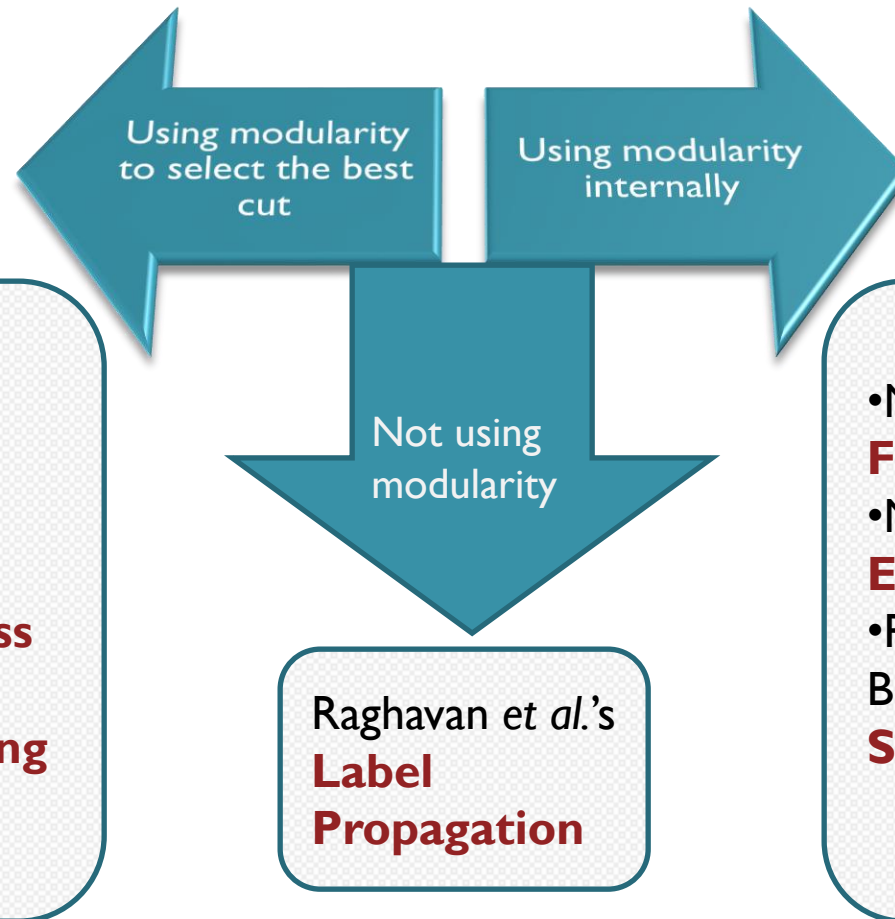
## Real Network Properties:

- Scale-free (power-law degree distribution)
- Power-law community size distribution
- Small average distance
- High transitivity
- High degree correlation





# 2.1 Selected Algorithms



## 2.2 Lancichinetti et. al generative model

$n, \beta, \gamma, \langle k \rangle, k_{max}, \mu$

● Apply configuration model with average degree  $\langle k \rangle$ , max. degree  $k_{max}$  and power-law exponent  $\gamma$

● Draw community sizes with power-law exponent  $\beta$  and affect each node to a community

● Rewire some links in order to respect  $\mu$ , without changing the nodes degrees

## 2.2 Lancichinetti *et. al* generative model

PARAMETER	VALUE
<b>n</b>	{100,500}
$\beta$	{1,2}
$\gamma$	{2,3}
$\langle k \rangle$	{5,15,30}
$k_{max}$	{15,45,90}
$\mu$	[0.05,0.95]

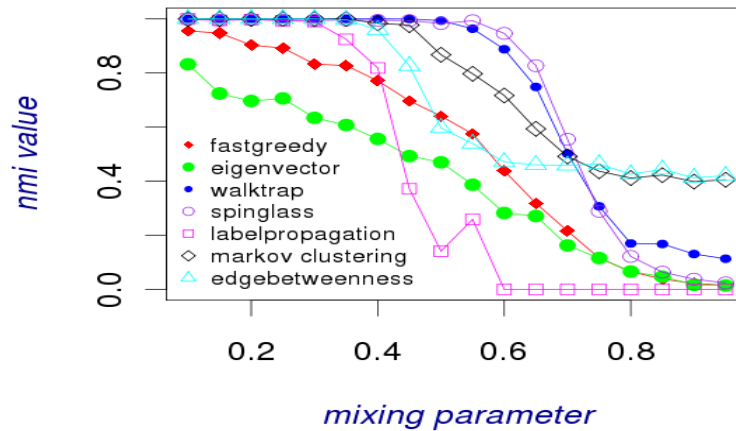
## 2.3 Normalized mutual information (NMI)

- Quality assessment for a dataset partition
  - $m$ : confusion matrix
  - $n$ : dataset size

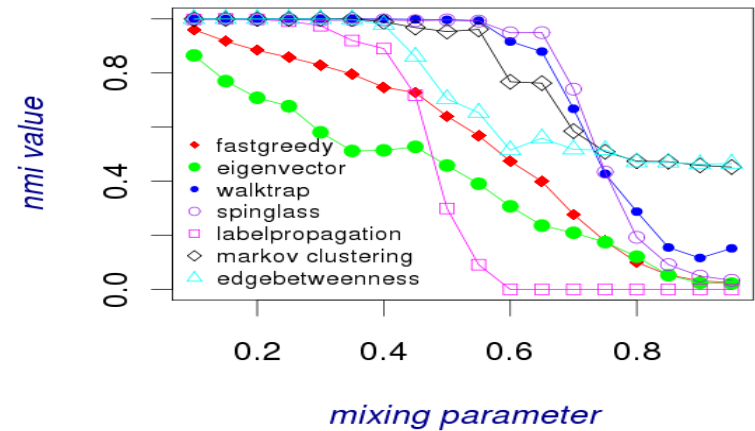
- Values:
$$I = \frac{-2 \sum_i \sum_j m_{ij} \log(nm_{ij}/m_{i+}m_{+j})}{\sum_i m_{i+} \log(m_{i+}/n) + \sum_j m_{+j} \log(m_{+j}/n)}$$
  - 0: random partition
  - 1: perfect partition

# 3. Results & Discussions

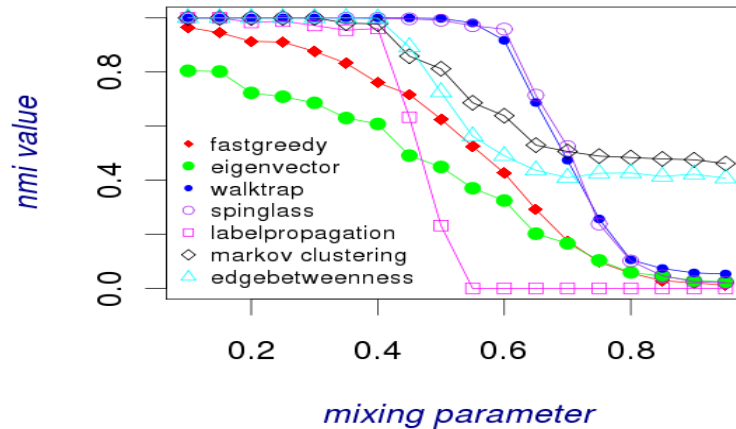
gamma 2 beta 1



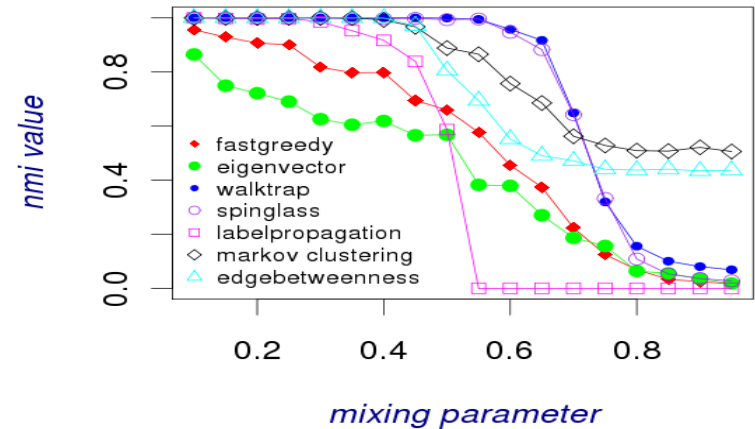
gamma 2 beta 2



gamma 3 beta 1



gamma 3 beta 2

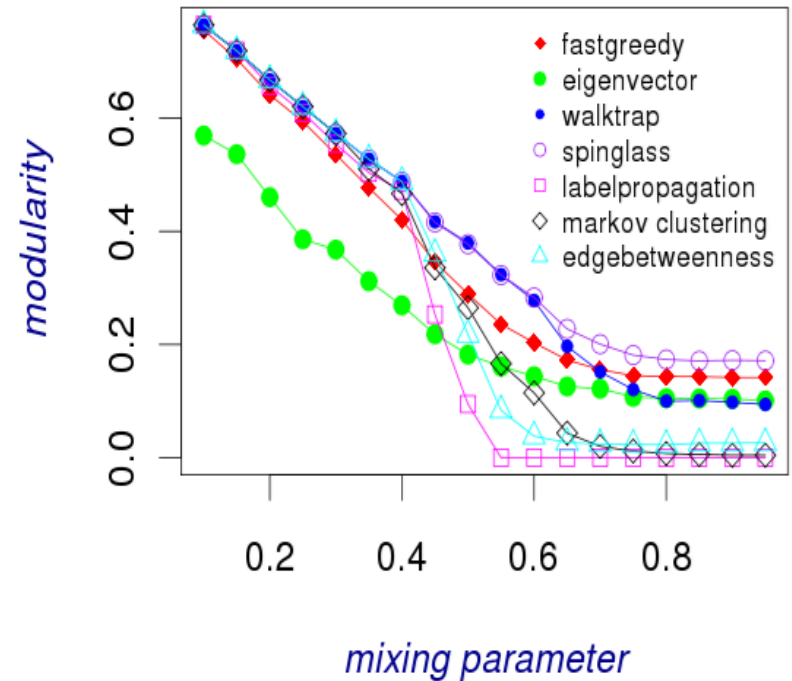
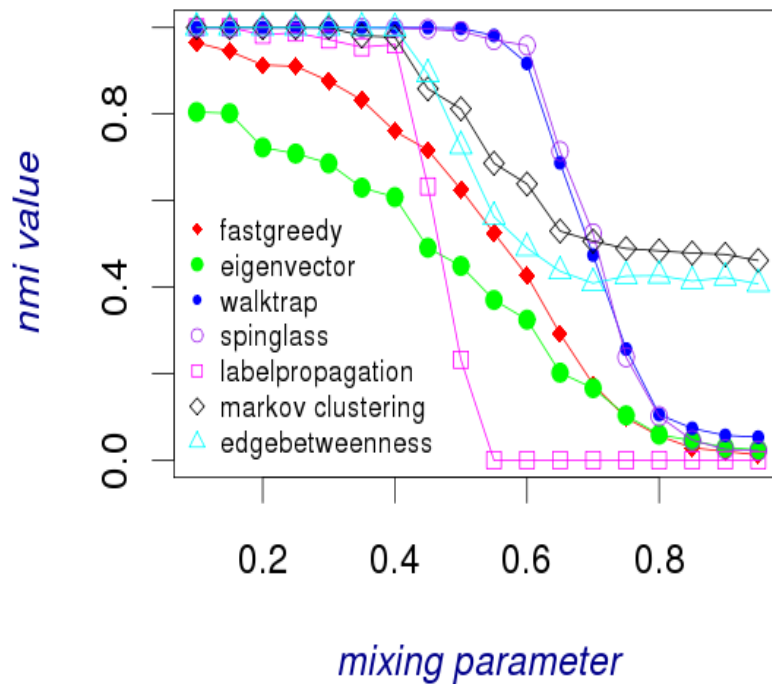


average degree 30

6/27/2009

# 3. Results & Discussions

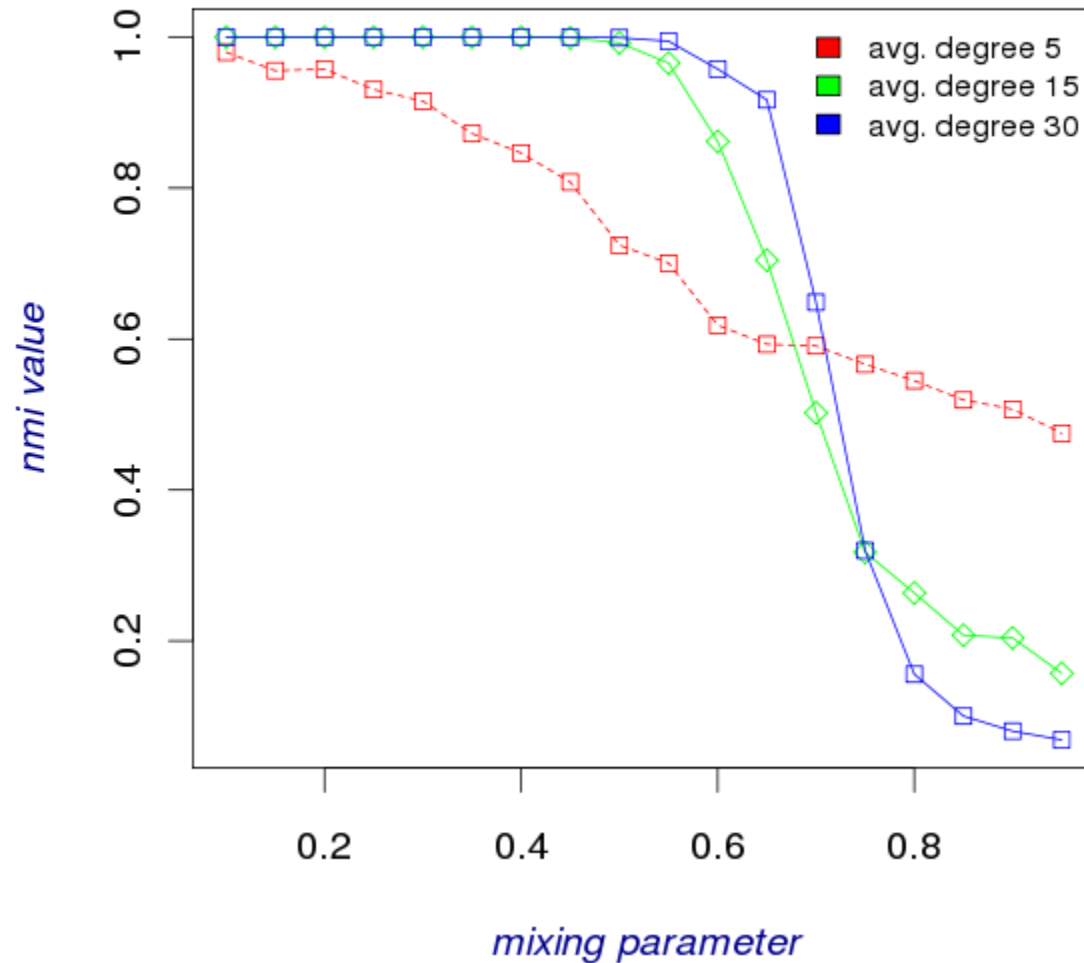
gamma 3 beta 1



average degree 30

# 3. Results & Discussions

WALKTRAP gamma 3 beta 2



# 3. Results & Discussions

- Parameters effects:
  - $\beta$  and  $\gamma$ : almost no effect ( $\rho < 0.06$ )
  - $\langle k \rangle$ : higher average degree improves performance
- Algorithms:
  - Partition quality:
    - MC, WT and SG performs better
    - EB and MC good for high  $\mu$
    - LP is not robust
  - Speed:
    - SG is slow, EB is very slow
    - LP, FG and WT are fast
    - EV and MC lie somewhere in between



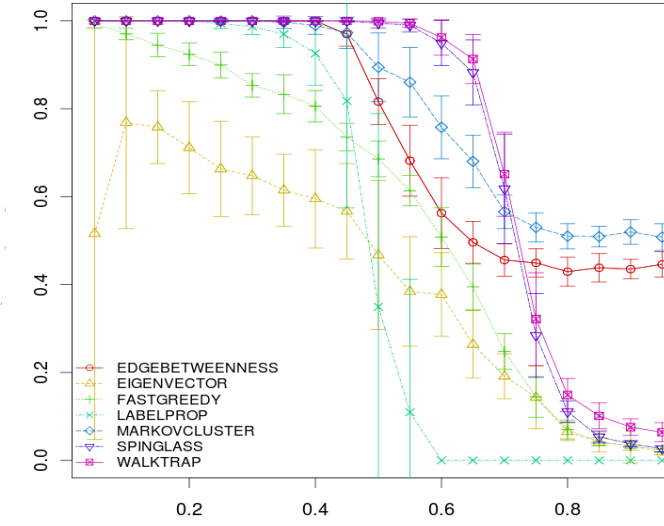
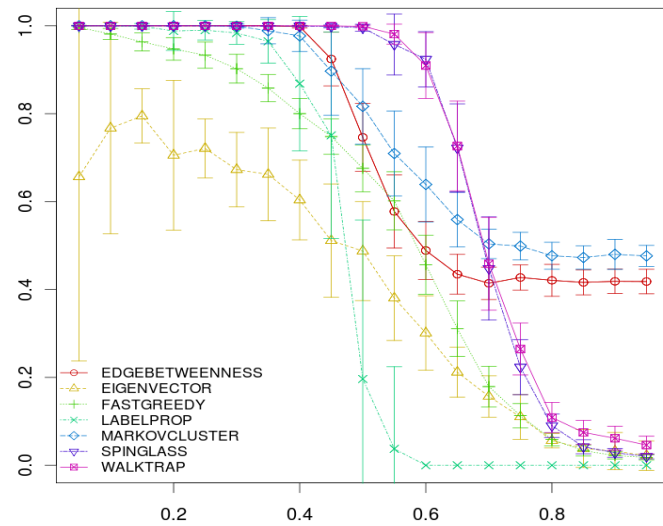
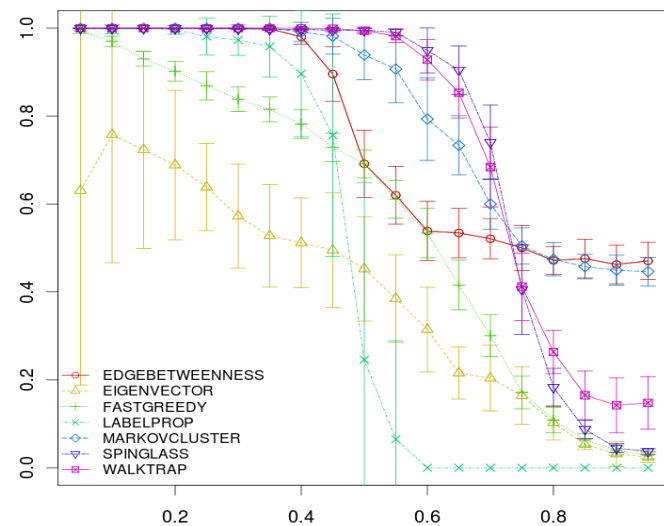
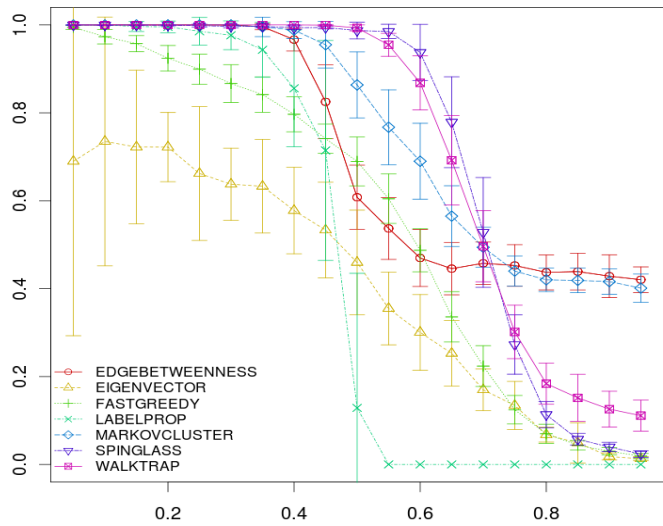
# 4. Conclusion

## Further experiments

- on larger networks
- with more realistic networks
- with more algorithms
- with different performance measures



# Questions



# Algorithms complexities

<b>LP</b>	<b>FG</b>	<b>EV</b>	<b>WT</b>	<b>MC</b>	<b>SG</b>	<b>EB</b>
$O(m+n)$	$O(n \log^2 n)$	$O(n^2)$	$O(n^2 \log n)$	$O(nk^2)$	<i>unknown</i>	$O(m^2 n)$

**n**: node number; **m**: link number ; **k**: number of ressource allocated

You can also give the complexities for sparse networks, they're simpler ( $m \approx n$ ). This k is weird, isn't it the average degree?

# Results

- Add a nmi plot with dispersion values, in case someone ask to see it
- Also, add a slide with only one NMI plot, in order to have something big to show the audience (in case they can't see what's on your 4-plots slides). And do the same with the corresponding modularity plot