



**HAL**  
open science

## Evaluating Community Detection Algorithms: A multidimensional issue

Chantal Cherifi, Hocine Cherifi

► **To cite this version:**

Chantal Cherifi, Hocine Cherifi. Evaluating Community Detection Algorithms: A multidimensional issue. International School and Conference on Network Science NETSCI 2018, Jun 2018, Paris, France. 2018. hal-01796539

**HAL Id: hal-01796539**

**<https://hal.science/hal-01796539v1>**

Submitted on 20 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

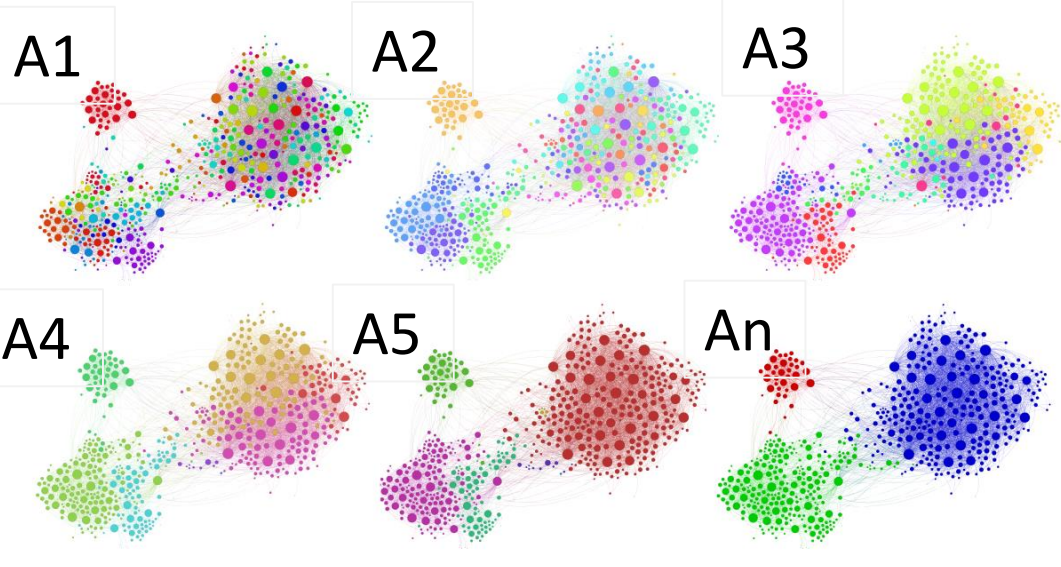


# Evaluating Community Detection Algorithms: A multidimensional issue

Chantal Cherifi  
University of Lyon 2 FRANCE

Hocine Cherifi  
University of Burgundy FRANCE

## Problematic



## Ranking the algorithms

**Hypotheses:**  
Ground-truth community structure available  
Overlapping community detection algorithms

## Classical approaches: Scalar

### Quality Desirable properties

	AD	AO	FO	ID	MO	OM
PGP	1.45	6.45	1.71	0.79	17.1	0.37
LFM	1.39	2.93	3	0.45	12.75	0.13
GCE	3.88	1.46	2.94	0.27	35.85	0.14
OSLOM	4.18	74.2	3.54	0.62	602.76	0.16

I: Average Degree (AD), Internal Density (ID), IE: Average ODF (AO), Flake ODF (FO), Max ODF (MO), M: Overlapping Modularity (OM)

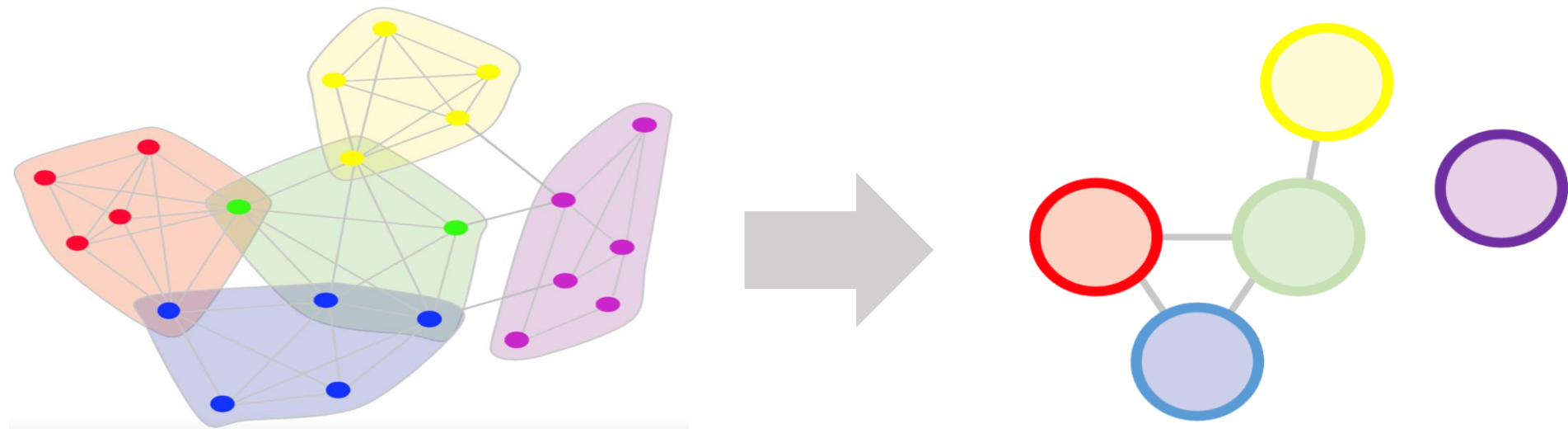
### Clustering Ground-truth comparison

	NMI	OI	F1-score
LFM	0.06	0.12	0.37
GCE	0.51	0.16	0.11
OSLOM	0.31	0.2	0.28

IT: Normalized Mutual Information (NMI), PC: Omega Index (OI), SC: F1-score

**Drawback: Same values can represent very different situations**

## Multidimensional approach: Set of topological properties of the community structure



Compare the ground-truth community structure with the uncovered community structure

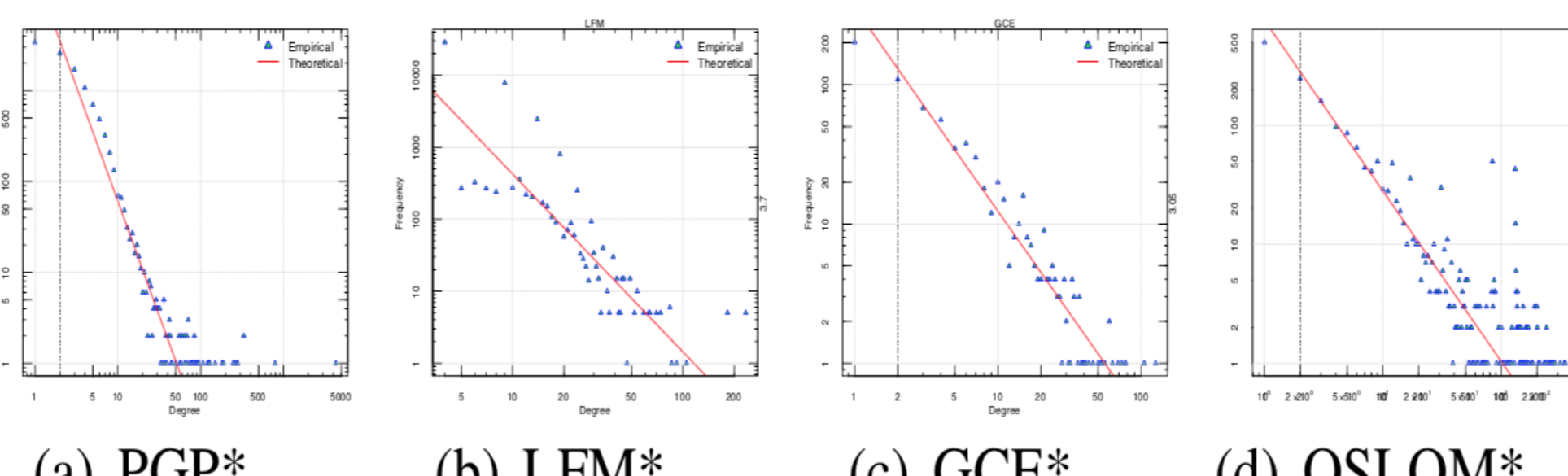
### Basic

	V	E	$\rho$	d	$l_G$	$\overline{deg}$	$\delta(G)$	$\tau$	C
PGP*	11074	23091	3.77E-04	15	7.43	4.17	4292	-0.12	0.01
LFM*	43558	146969	1.55E-04	26	9.12	6.75	234	0.15	0.61
GCE*	741	2840	1.04E-02	10	5.77	7.67	126	-0.02	0.2
OSLOM*	1972	22778	1.17E-02	10	4.1	23.1	348	0.21	0.64

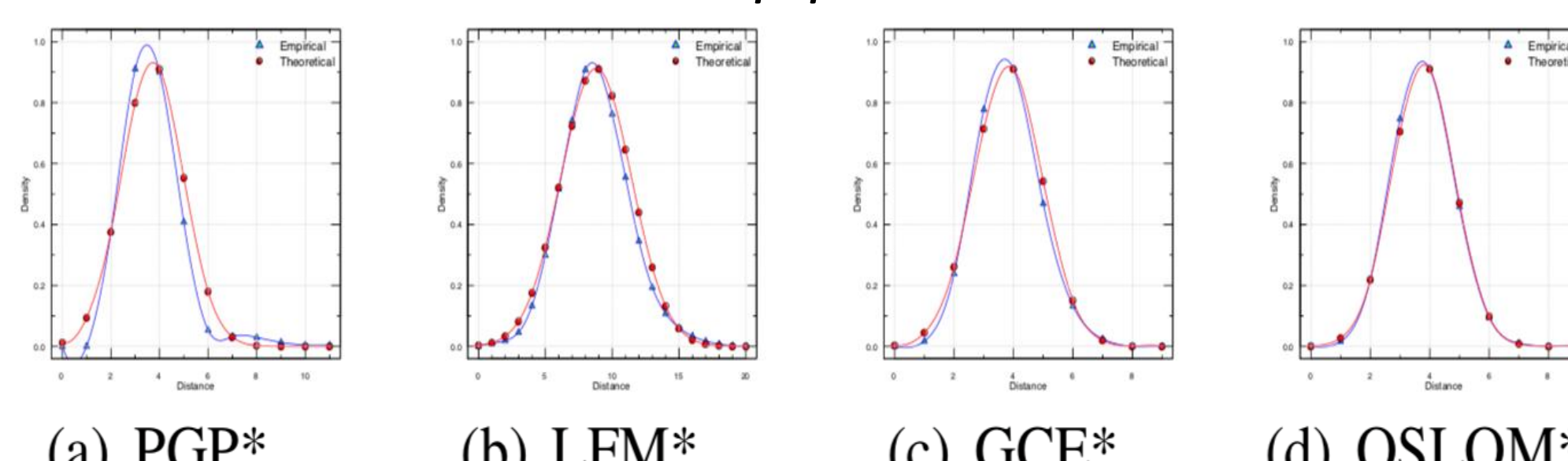
Number of nodes (V), Number of edges (E), Density ( $\rho$ ), Diameter (d), Average shortest path ( $l_G$ ), Average node degree ( $\overline{deg}$ ), Max node degree ( $\delta(G)$ ), Assortativity ( $\tau$ ), Clustering (C)

### Microscopic

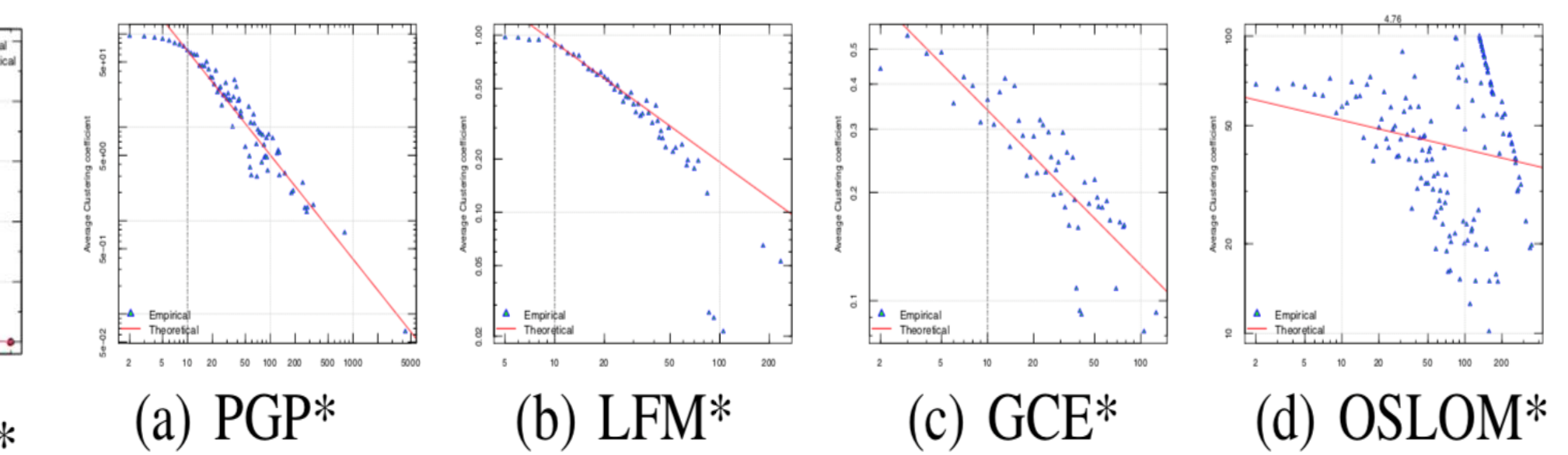
#### Degree distribution



#### Hop-plot

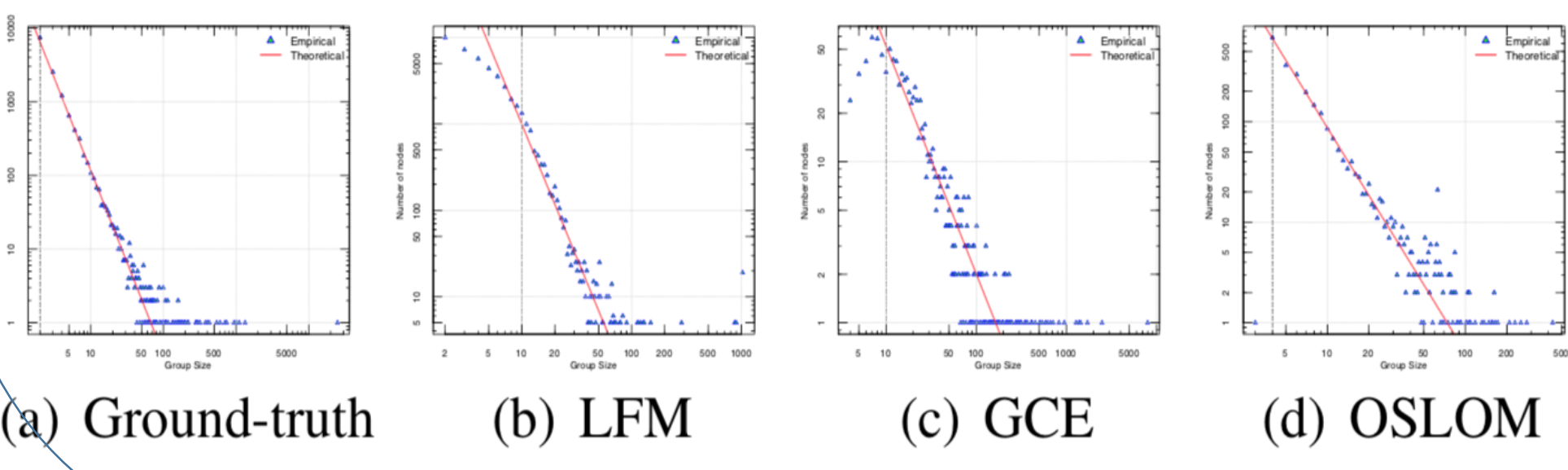


#### Average clustering as a function of node degree

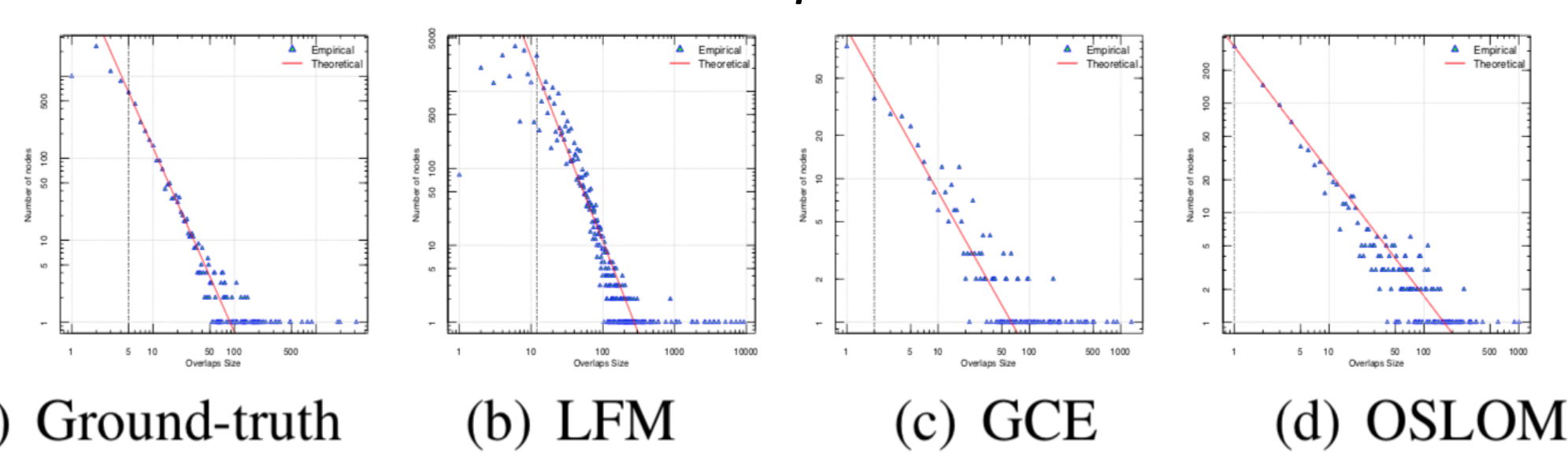


### Mesoscopic

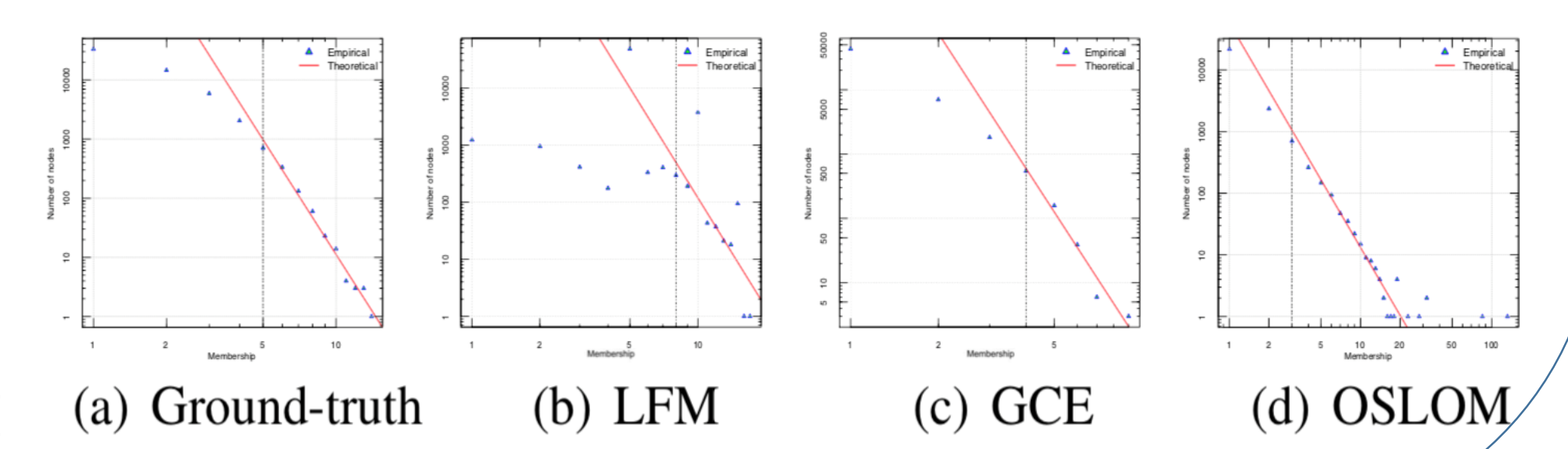
#### Community size



#### Overlap size

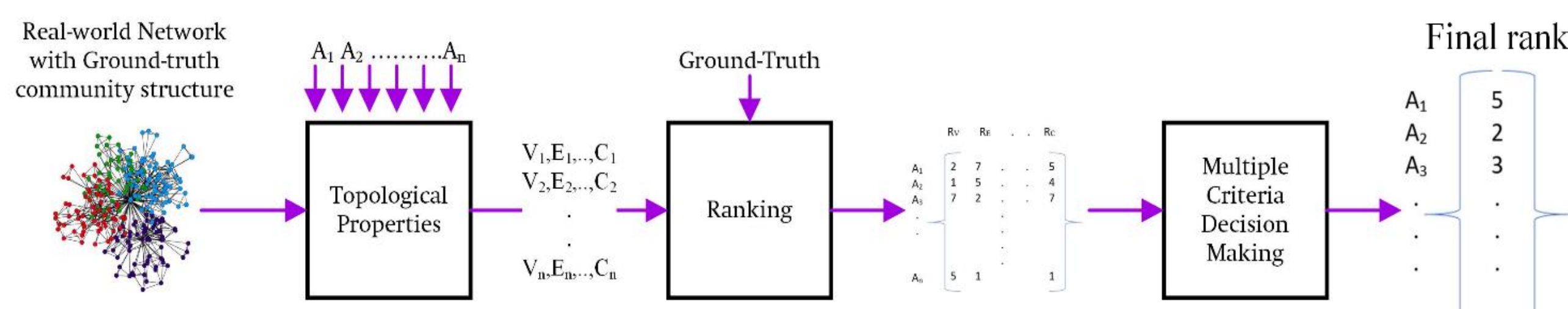


#### Node membership

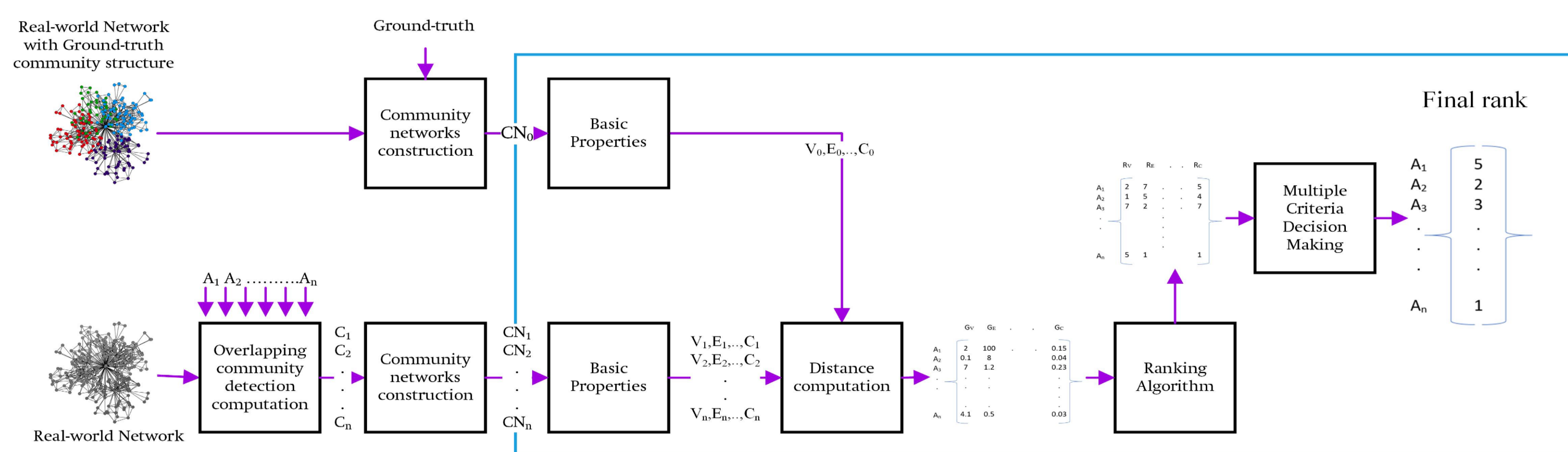


## Ranking the algorithms

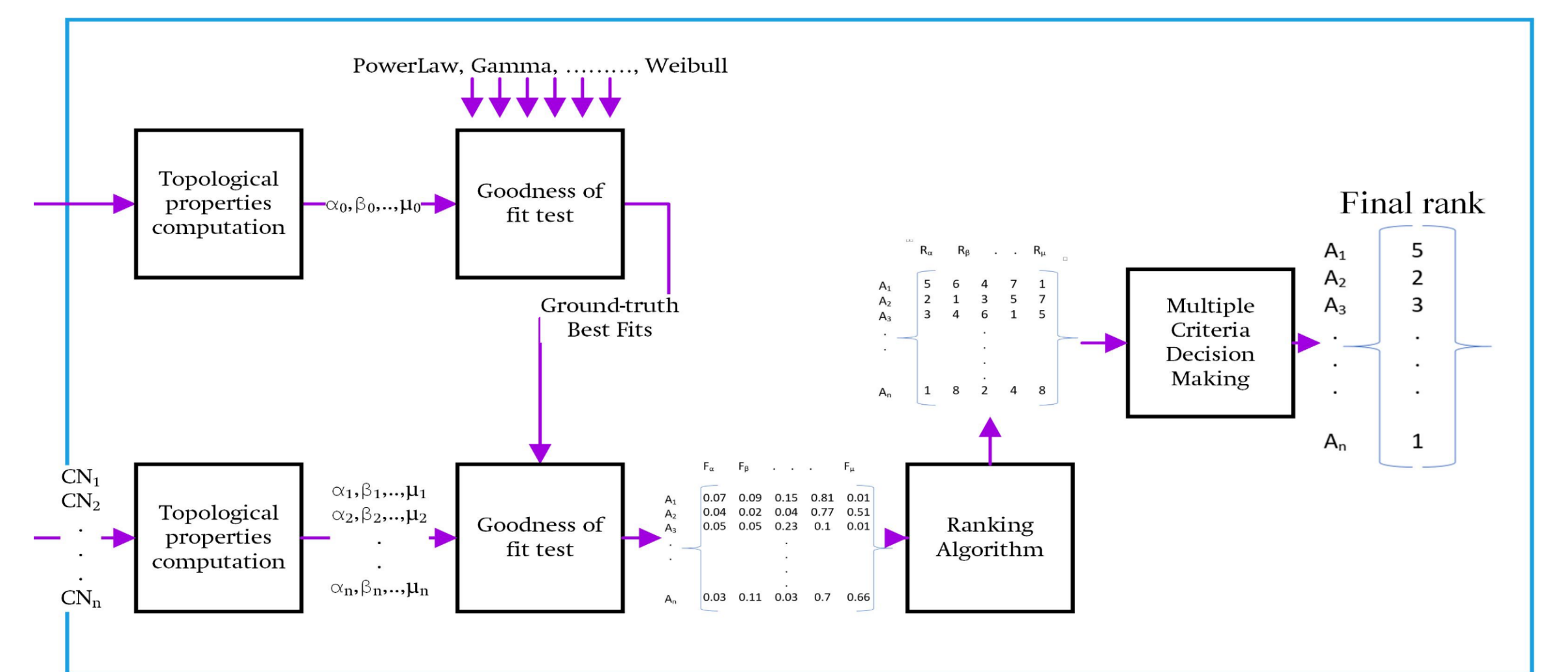
### General scheme



### Scalar



### Distribution



KS value degree distribution ground-truth: Power-Law (PL), Beta (BE), Cauchy (CA), Exponential (E), Gamma (GM), Logistic (LO), Log-Normal (LN), Normal (N), Uniform (U), Weibull (WB)

## Results

### Quality

	AD	AO	FO	ID	MO	OM
AD	1					
AO	0.29	1				
FO	0.54	-0.43	1			
ID	-0.04	0.11	-0.29	1		
MO	0.89	0.43	0.57	0.04	1	
OM	0	0.54	-0.32	0.25	0.04	1

Redundancy? →

### Clustering

	NMI	OI	F1-score
NMI	1		
OI	0.42	1	
F1-score	0.35	0.71	1

### Community structure (Topo)

	Basic	Micro	Meso
Basic	1		
Micro	0.61	1	
Meso	0.11	0.28	1

→ Low correlation

Each dimension shows a different view of the community structure

Topo, Quality or Clustering? →

	Topo	Quality	Clustering
Topo	1		
Quality	0.96	1	
Clustering	0.57	0.43	1

	Topo	Quality	Clustering
Topo	1		
Quality	0.6	1	
Clustering	0.82	0.32	1

→ Topo is more flexible

The winner? →

	Basic properties									Microscopic			Mesoscopic			Clustering						Quality			MCDM Ranking	
	V	E	$\rho$	d	$l_G$	$\overline{deg}$	$\delta(G)$	$\tau$	C	DD	Av	HD	CS	M	OS	AD	AO	FO	ID	MO	OM	NMI	OI	F1-score	Kconsensus	TOPSIS
LFM	7	6	1	7	4	3	5	6	5	5	1	4	6	6	1	1	4	3	5	2	7	7	7	4	5	3
GCE	4	5	5	3	3	4	7	3	3	3	5	1	3	7	6	5	5	2	7	4	6	3	6	7	7	5
OSLOM	3	1	6	3	7	5	4	7	6	6	6	7	4	5	5	6	7	5	2	7	5	4	5	5	4	6
LINKC	6	7	2	5	1	7	3	4	7	2	7	8	7	3	3	4	2	6	6	5	1	5	1	2	3	4
SVINET	1	2	4	1	5	2	1	1	1	1	2	2	2	2	4	2	1	4	1	1	2	1	2	1	1	1
SLPA	2	4	3	2	6	1	2	2	2	4	4	6	1	1	2	3	6	1	4	3	3	2	3	3	2	2
DEMON	5	3	7	6	2	6	6	4	4	7	3	5	5	4	7	7	3	7	3	6	4	6	4	6	6	7

## Open questions

1. What is the space size?
2. Ground-truth unavailable?
3. Non-overlapping communities?