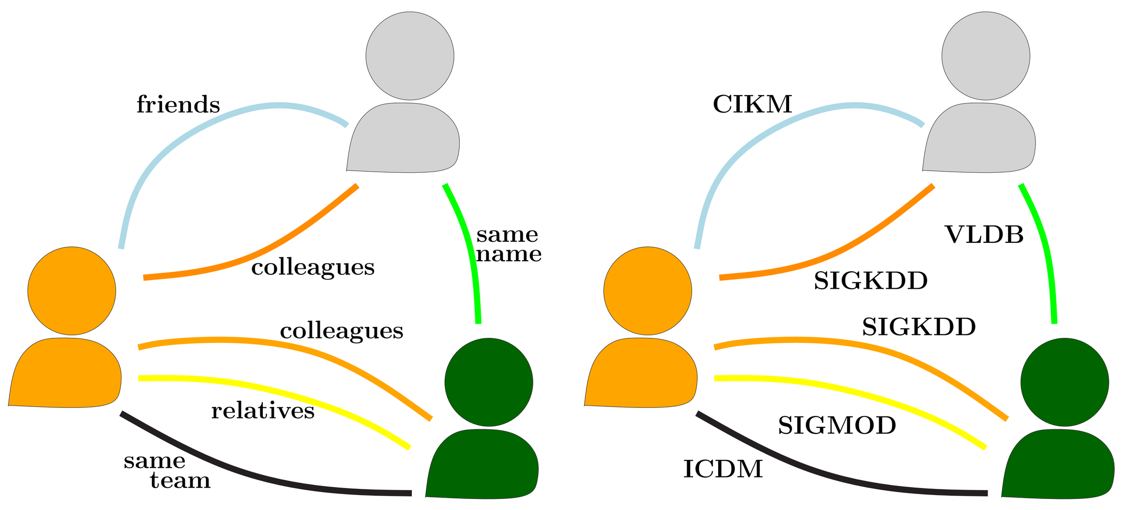
Finding Redundant and Complementary Communities in Multidimensional Networks

Michele Berlingerio, Michele Coscia, Fosca Giannotti
KDDLab, ISTI CNR, Pisa, Italy

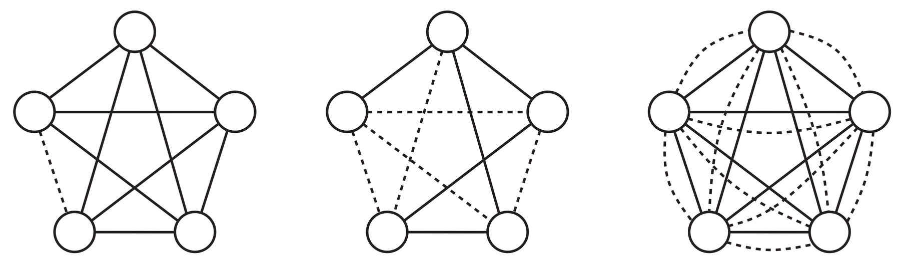
Problem



Our setting: Multidimensional Networks

Problem 1 ($\mathcal{MCD}$) *Given a multidimensional network $\mathcal{G}$, find and characterize the multidimensional communities.*

Characterization



(a) (b) (c)

Examples of multidimensional communities

Characterization of a community c via γ (*complementarity*):

- **Variety** $\mathcal{V}_c$: # of different dimensions
- **Exclusivity** $\mathcal{E}_c$: # of pairs connected by only one dimension
- **Homogeneity** $\mathcal{H}_c$: distribution of edges over dimensions

We aggregate the above by their product:

$$\gamma_c = \mathcal{V}_c \times \mathcal{E}_c \times \mathcal{H}_c \quad (1)$$

Characterization of c via ρ (*redundancy*):

- **Redundancy** ρ_c: level of replication of pairwise connections between any two nodes

We formulate the above as:

$$\rho_c = \sum_{(u,v)\in \overline{\overline{P_c}}} \frac{|\{d : \exists(u,v,d) \in E\}|}{|D| \times |P_c|} \quad (2)$$

Characterization of the above three examples:

(a) γ=0, ρ=0
(b) γ=1, ρ=0
(c) γ=0, ρ=1

References

[1] A. Clauset, M. E. J. Newman and C. Moore, Finding community structure in very large networks. In *Physical Review E*.
[2] S. Fortunato, Community detection in graphs In *Physics Reports*.
[3] J. Leskovec, K. J. Lang and M. W. Mahoney, Empirical comparison of algorithms (...). In *WWW 2010*.
[4] P. Pons and M. Latapy, Computing Communities in Large Networks Using Random Walks. In *ISCIS 2005*.
[5] U.N. Raghavan, R. Albert and S. Kumara, Near linear time algorithm (...). In *Physical Review E*.

Algorithm

Algorithm 1 *MCD_Solver*

Require: $\mathcal{G}, \phi, CD$
Ensure: set of multidimensional communities $\mathcal{C}$ and sets of their charaterization S_γ, S_ρ

```
1: G ← φ(𝒢)
2: C ← CD(G)
3: for all c′ ∈ C do
4:    c ← φ′(c′)
5:    𝒞 ← 𝒞 ∪ c
6:    S_ρ ← S_ρ ∪ ρ(c)
7:    S_γ ← S_γ ∪ γ(c)
8: end for
9: return 𝒞, Γ, P
```

Mapping Function ϕ

We use three different ϕ:

Connectivity check:

$$\mu_{u,v} = \begin{cases} 1 & \text{if } \{\exists\, d : (u,v,d) \in E\} \\ 0 & \text{otherwise} \end{cases}$$

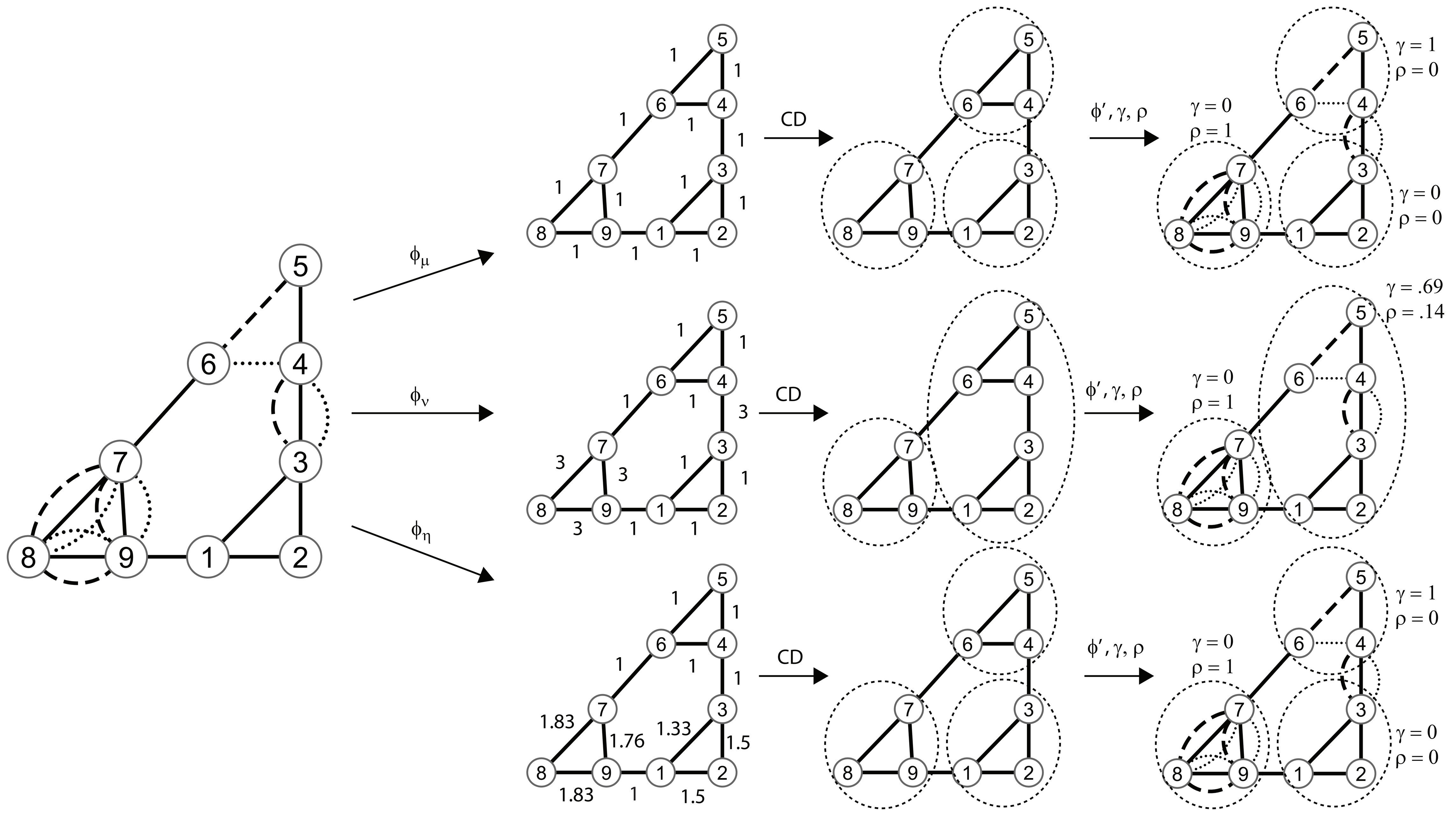
Number of dimensions:

$$\nu_{u,v} = |\{d : (u,v,d) \in E\}|$$

Multidimensional clustering coefficient:

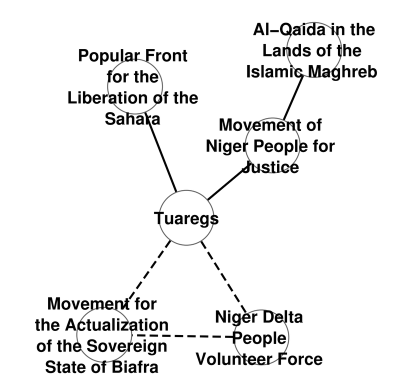
$$\eta_{u,v} = 1 + \frac{|N_{u,l} \cap N_{v,l}|}{|N_{u,l} \cup N_{v,l}| - 2}$$

Run Through Example

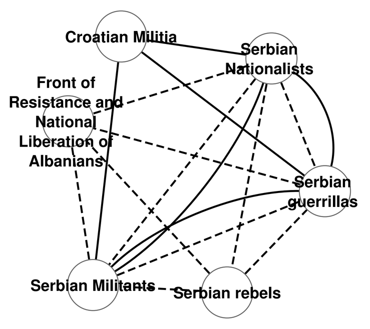


Run through example for three instances of MCD_Solver varying the ϕ parameter

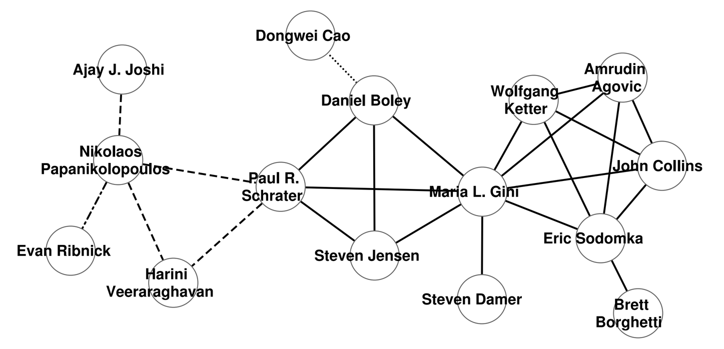
Interesting Communities



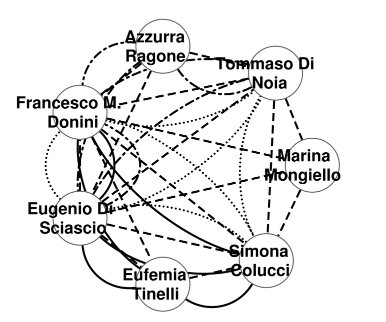
(a) high γ (0.0081) in GTD



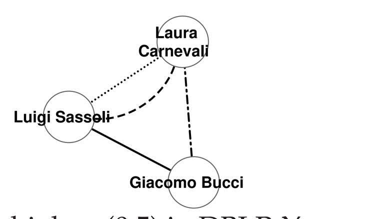
(b) high ρ (0.0035) in GTD



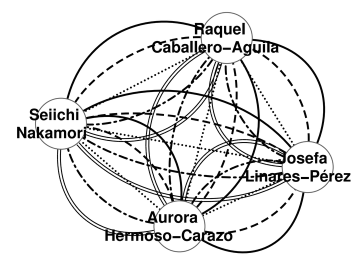
(c) high γ (0.00064) in DBLP-C



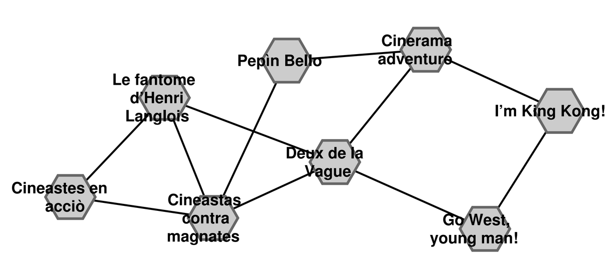
(d) high ρ (0.06048) in DBLP-C



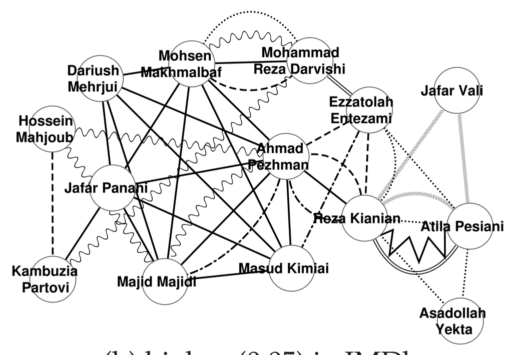
(e) high γ (0.5) in DBLP-Y



(f) high ρ (1) in DBLP-Y

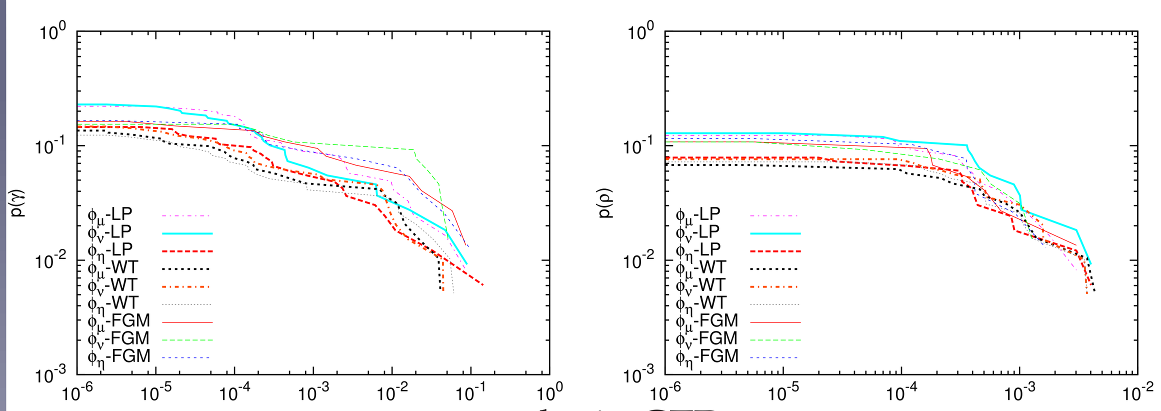


(g) high γ (0.005) in IMDb



(h) high ρ (0.05) in IMDb

Quantitative Evaluation



γ and ρ in GTD

Network	ϕ	LP $\lvert\mathcal{C}\rvert$	LP Q	WT $\lvert\mathcal{C}\rvert$	WT Q	FGM $\lvert\mathcal{C}\rvert$	FGM Q
GTD	$\phi\mu$	122	**0.622**	192	0.620	74	0.584
	$\phi\nu$	109	0.547	197	0.603	65	0.611
	$\phi\eta$	165	0.500	194	**0.621**	78	**0.616**
DBLP-C	$\phi\mu$	4625	**0.793**	4216	0.819	2931	0.860
	$\phi\nu$	4685	0.791	4629	0.810	2820	0.881
	$\phi\eta$	5983	0.783	4345	**0.837**	2869	**0.903**
DBLP-Y	$\phi\mu$	1632	0.190	5064	0.561	980	0.593
	$\phi\nu$	8088	**0.591**	6397	0.638	754	**0.730**
	$\phi\eta$	8084	0.584	6131	**0.643**	722	0.723
IMDb	$\phi\mu$	87	0.415	860	0.494	64	0.442
	$\phi\nu$	124	**0.483**	847	**0.541**	66	**0.536**
	$\phi\eta$	148	0.460	823	0.507	63	0.530

#communities ($|\mathcal{C}|$) and modularity (Q)

Contacts

The authors may be contacted at:
{name.surname}@isti.cnr.it
http://kdd.isti.cnr.it