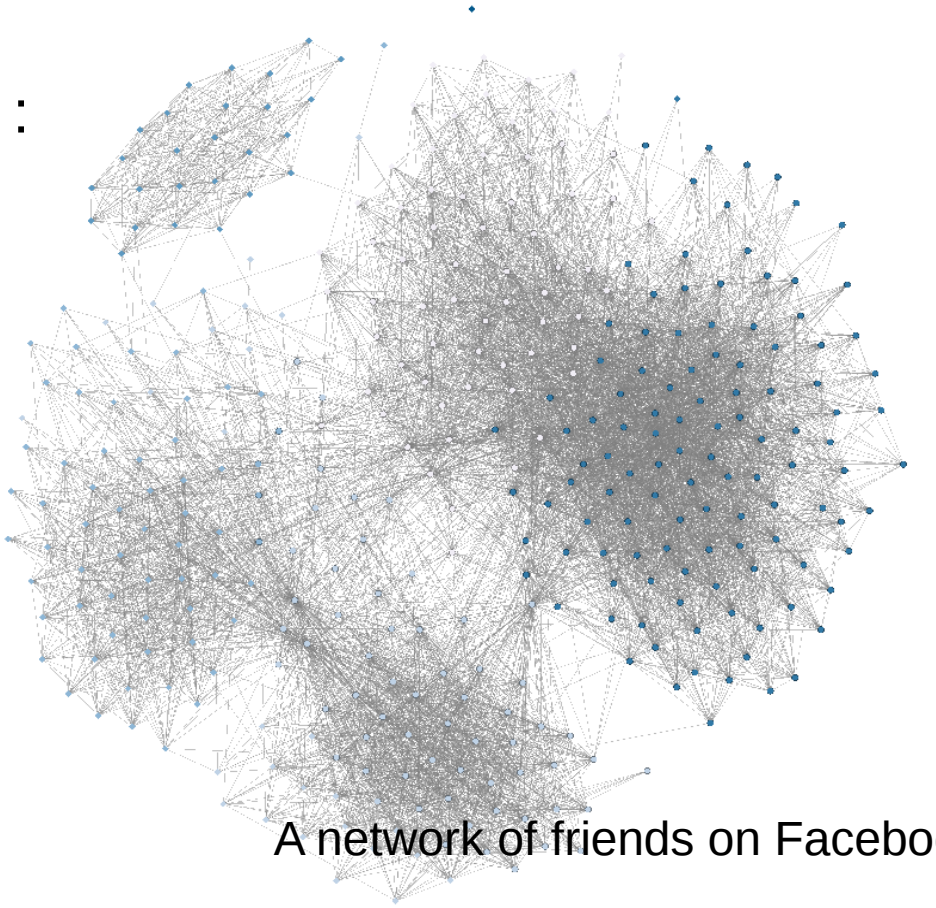
Revealing properties of the bipartite structure of online social networks

Raphaël TACKX,
Jean-loup GUILLAUME,
Fabien TARISSAN

Context

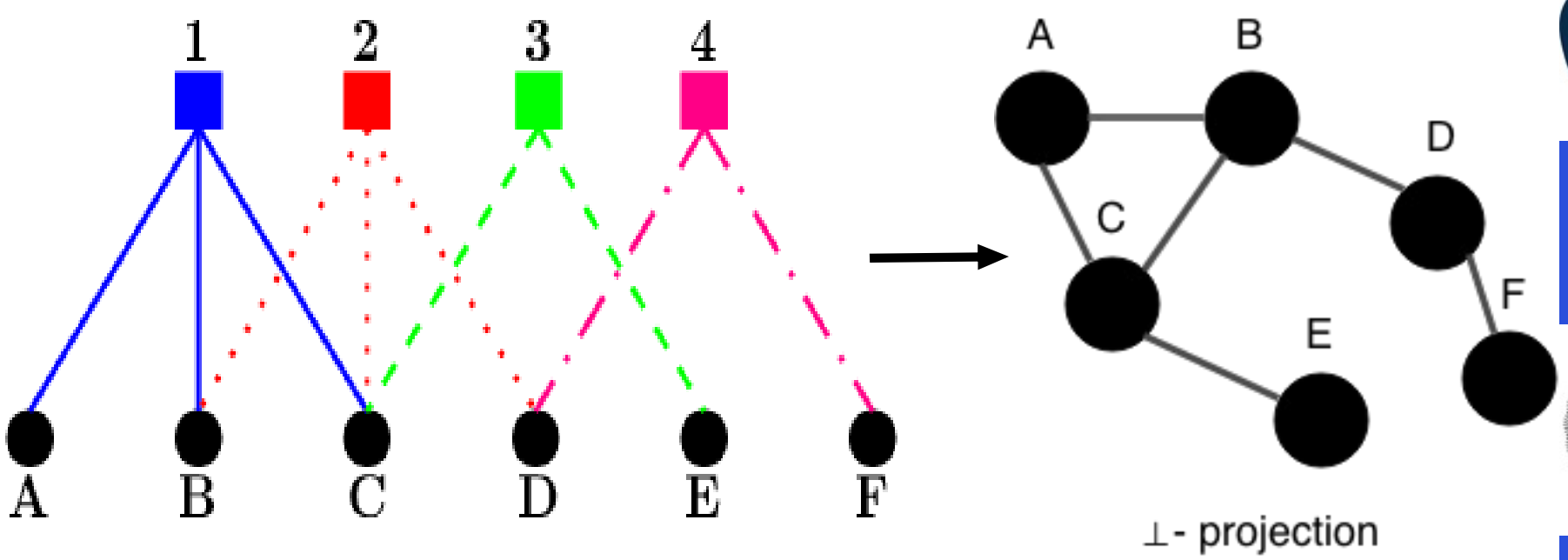
- Complex networks studies :
  - Social networks
  - Biology
  - Neuron networks
  - Economy
- Fundamental properties :
  - Weak global density
  - Strong clustering
  - Average short distance
  - Heterogeneous degree **distribution**
  - Existence of **communities**
  - Assortativity or disassortativity
- Why study **real-world** networks ?



A network of friends on Facebook

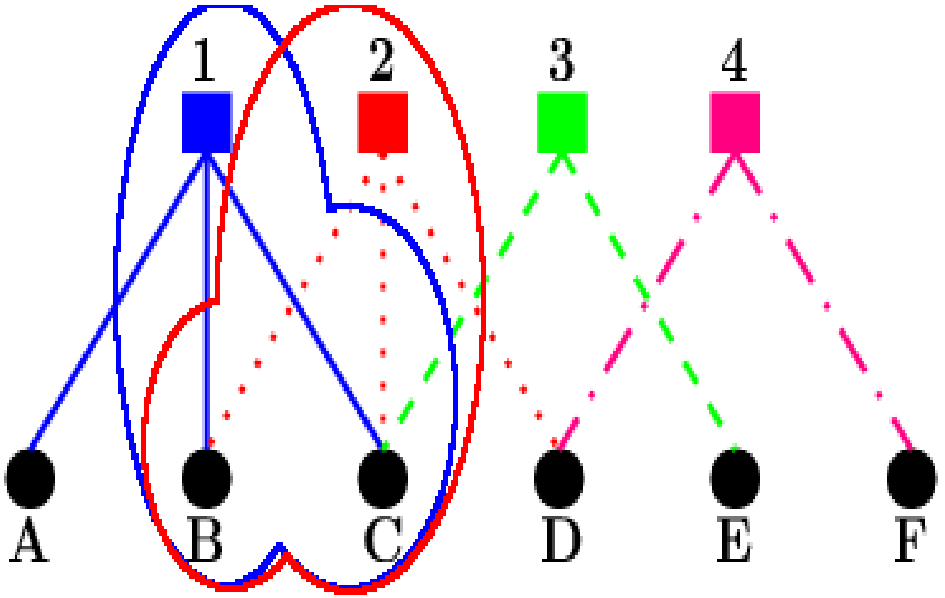
Unipartite VS bipartite graphs

- Some networks have a natural bipartite organisation
  - Movies-Actors
  - Peer-to-peer (Files-Users)
  - Authors-Publications

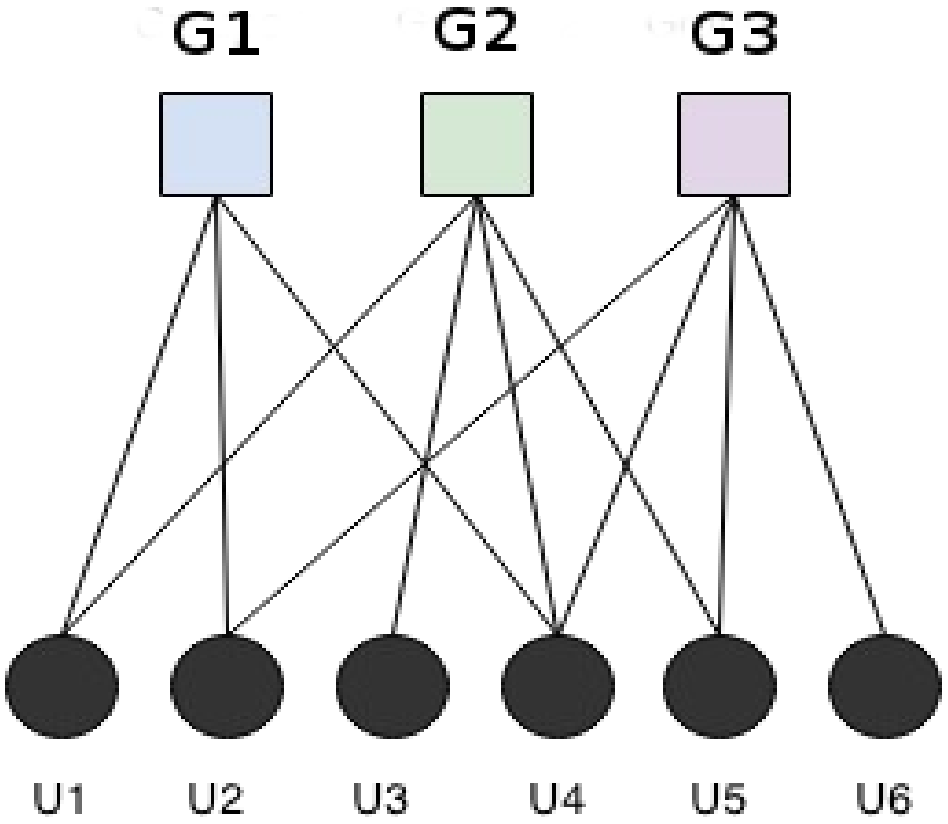


Our motivation

- Study of **bipartite real-world networks**
  - Extract fundamental properties
  - Better understand the structure and the features
- One of the issues : the overlaps



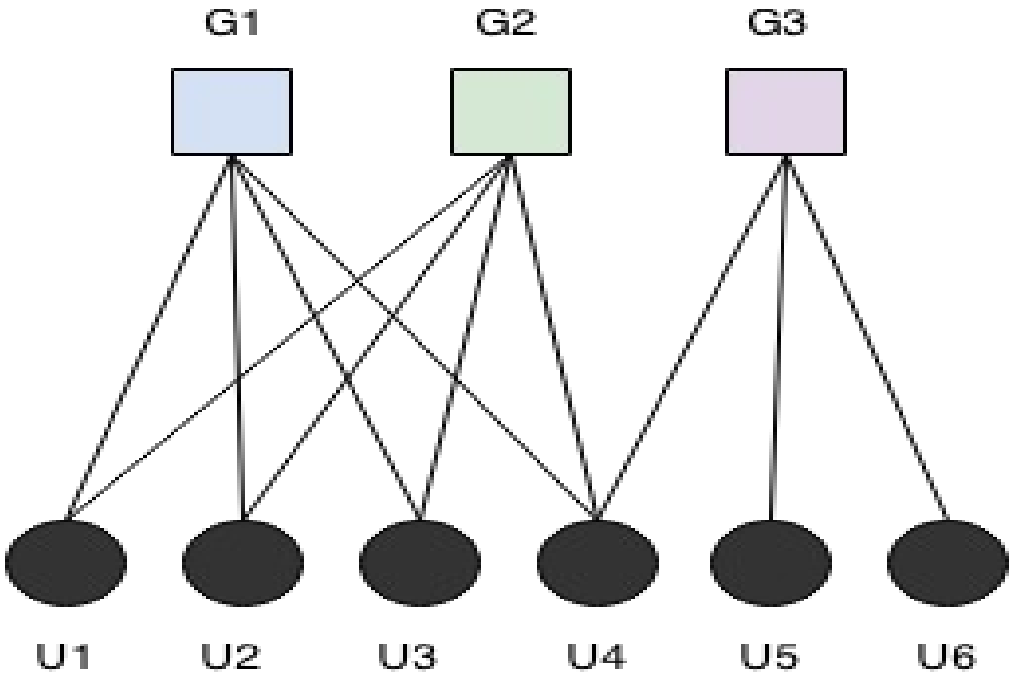
Datasets



Example : **LiveJournal** dataset
(groups / users)

Metrics

- Clustering coefficient : $cc_{\top}(u,v) = \frac{|N_{\top}(u) \cap N_{\top}(v)|}{|N_{\top}(u) \cup N_{\top}(v)|}$



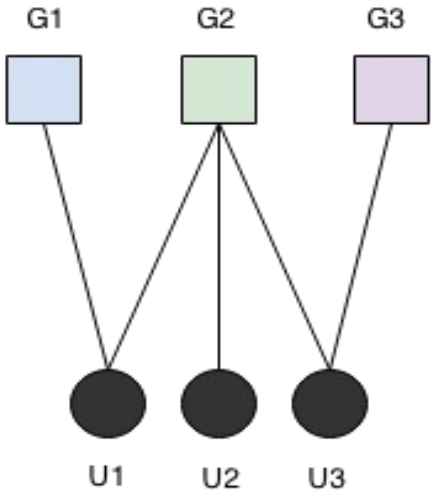
$$cc_{\top}(G2,G1) = \frac{|\{U1,U2,U3,U4\}|}{|\{U1,U2,U3,U4\}|} = 1$$

$$cc_{\top}(G2,G3) = \frac{|\{U4\}|}{|\{U1,U2,U3,U4,U5,U6\}|} = \frac{1}{6}$$

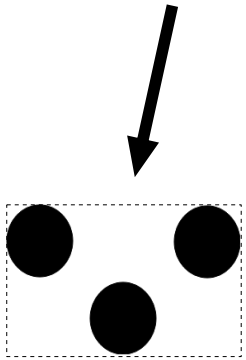
Metrics

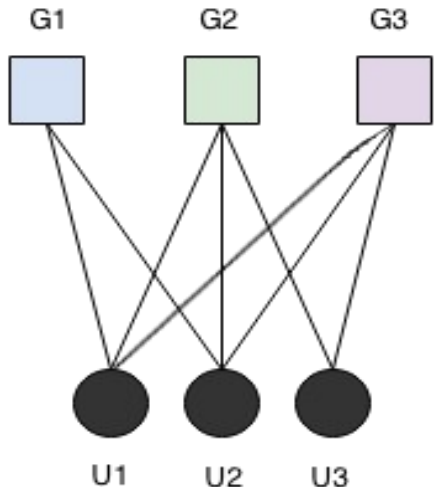
- Redundancy coefficient :

$$rc_{\top}(u) = \frac{|\{\{v,w\} \in N_{\top}(u)^2, \exists u' \neq u, (u',v) \in E, (u',w) \in E\}|}{\frac{|N_{\top}(u)| \cdot (|N_{\top}(u)|-1)}{2}}$$

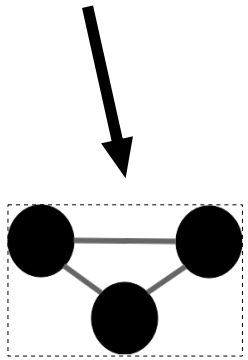


$rc_{\top}(G2) = 0$



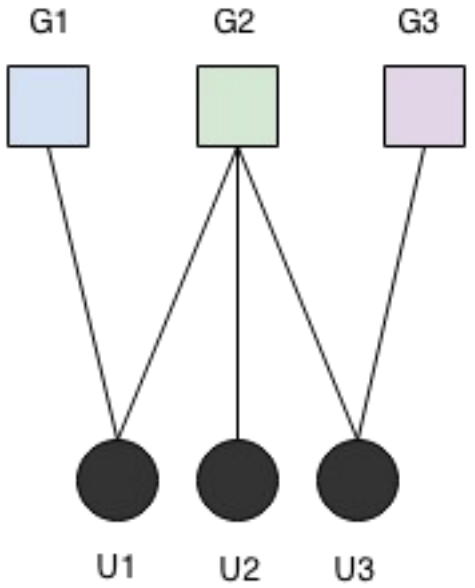


$rc_{\top}(G2) = 1$

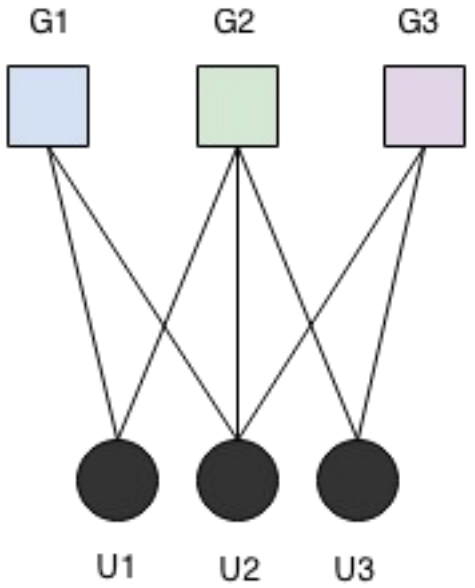


New metrics

- Dispersion : $$dispersion_{\top}(u) = \frac{|N_{\top}(N_{\perp}(u))|}{\sum_{v \in N_{\perp}(u)} (|N_{\top}(v)| - 1)}$$



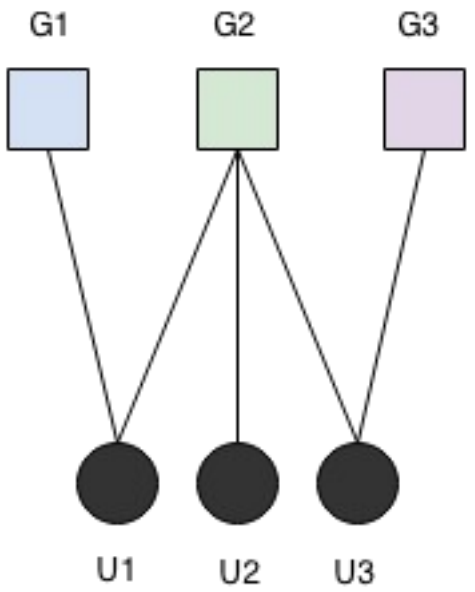
$$dispersion_{\top}(G2) = \frac{|\{G1, G3\}|}{2} = 1$$



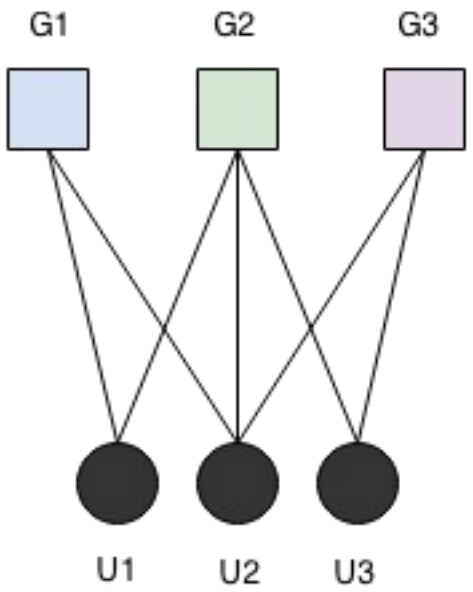
$$dispersion_{\top}(G2) = \frac{|\{G1, G3\}|}{4} = \frac{1}{2}$$

New metrics

- Monopoly : $$monopoly_{\top}(u) = \frac{\sum_{v \in N_{\perp}(u)} |N_{\top}(v)| = 1}{|N(u)|}$$



$$monopole_{\top}(G2) = \frac{1}{3}$$



$$monopole_{\top}(G2) = 0$$

Our approach

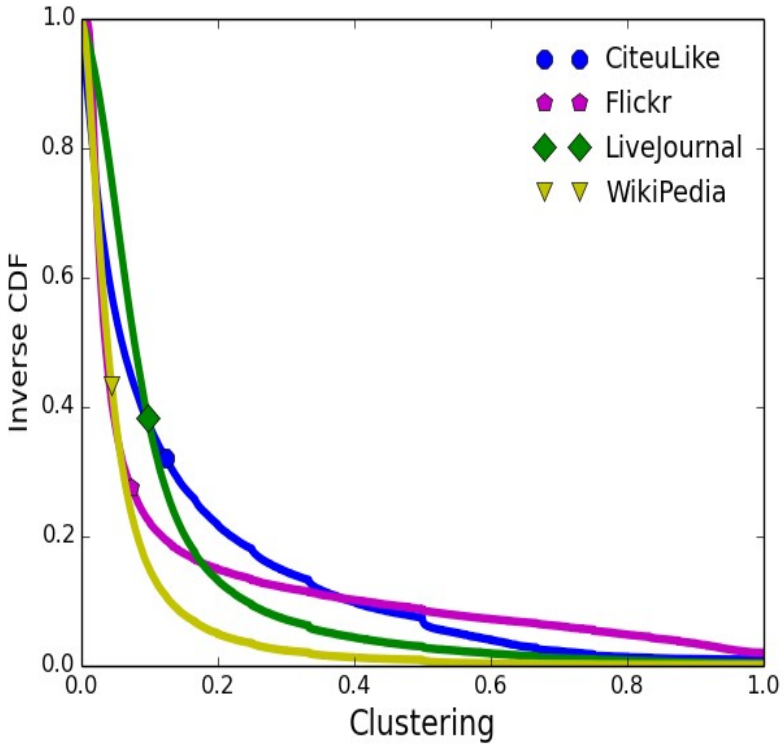
- Empirical studies :
  - CiteULike : tags / publications
  - LiveJournal : groups / users
  - Wikipedia : categories / articles
  - Flickr : groups / users
- Nearly the same size :
  - ~ 350 000 $\top$ nodes (top)
  - ~ 1,5 million of $\bot$ nodes (bottom)
- Questions :
  - **Comparison between CC and RD**
  - **Contribution of the new metrics**
  - Comparison with a random model

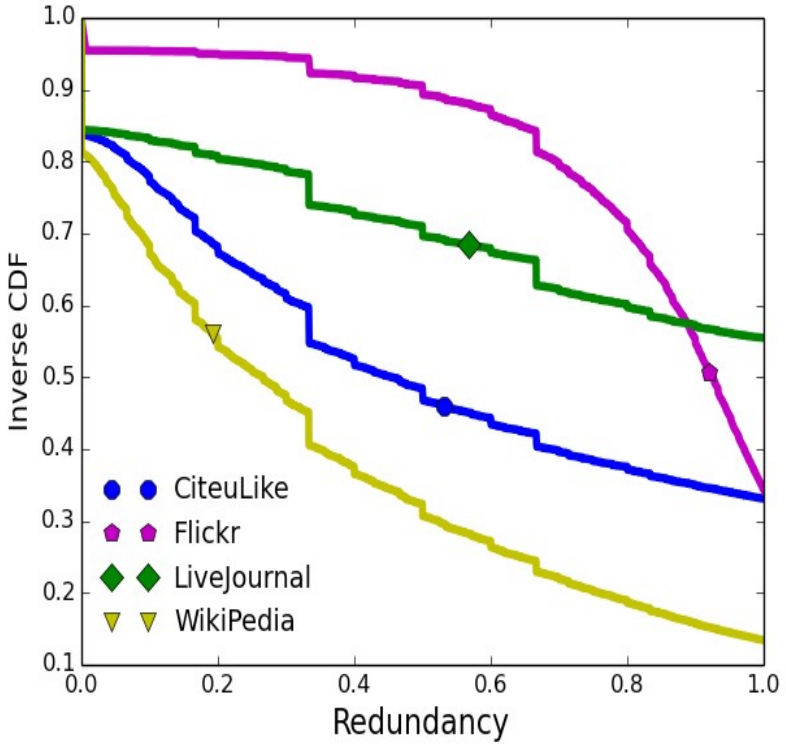
Results

- Average values of CC and RD

	CiteULike	LiveJournal	WikiPedia	Flickr
cc_{bip}	0.138	0.117	0.063	0.055
rd_{bip}	0.521	0.703	0.387	0.646

- Distribution of CC and RD

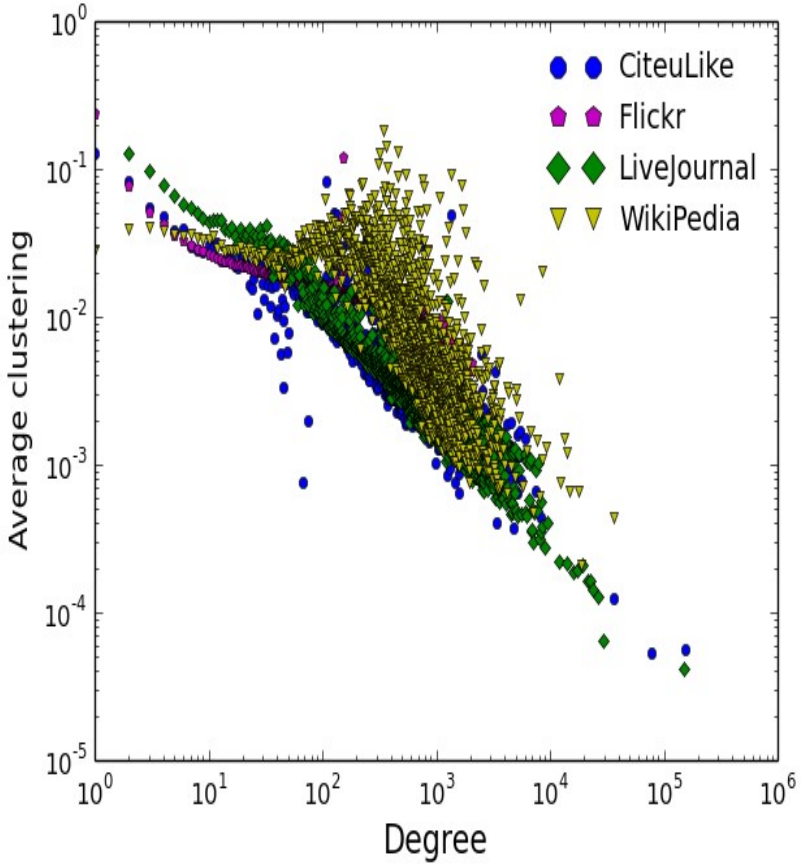


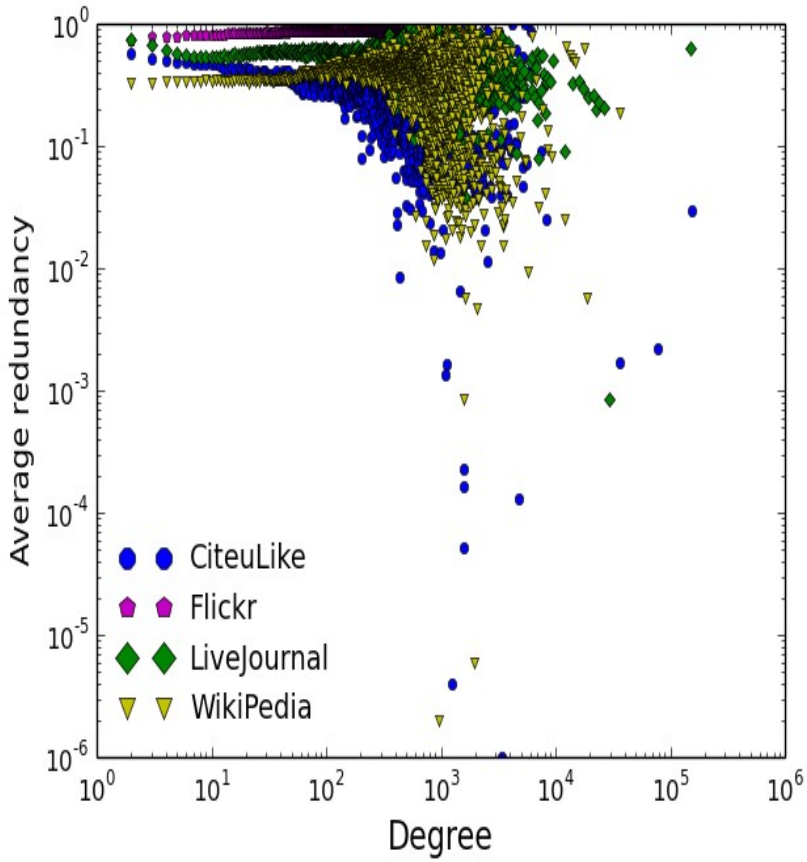


Important differences between both metrics

Results

- Correlation between degree/CC and degree/RD

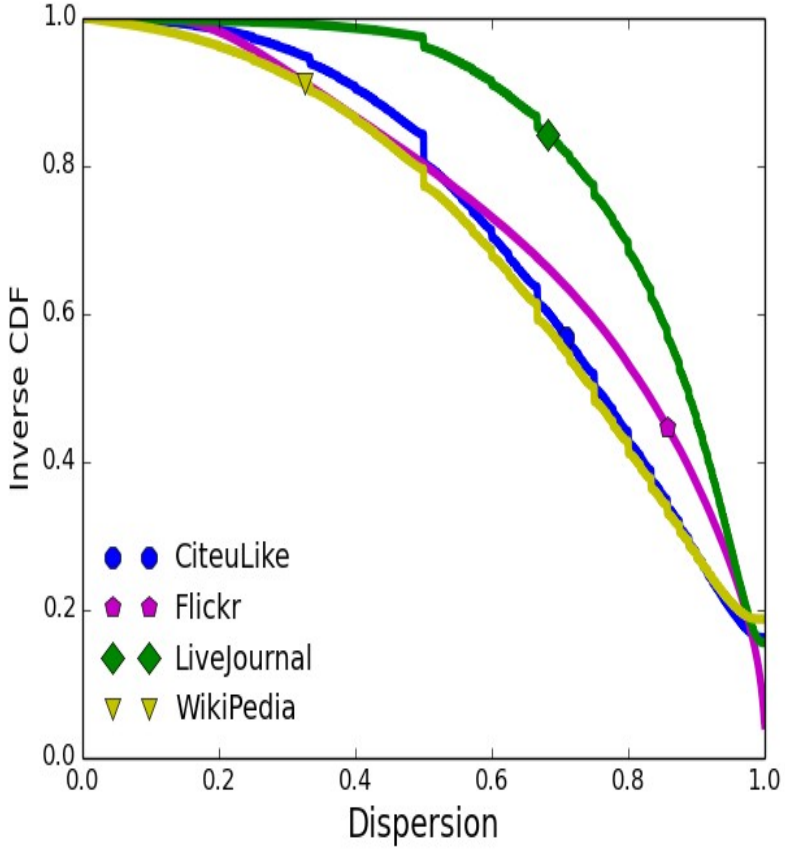




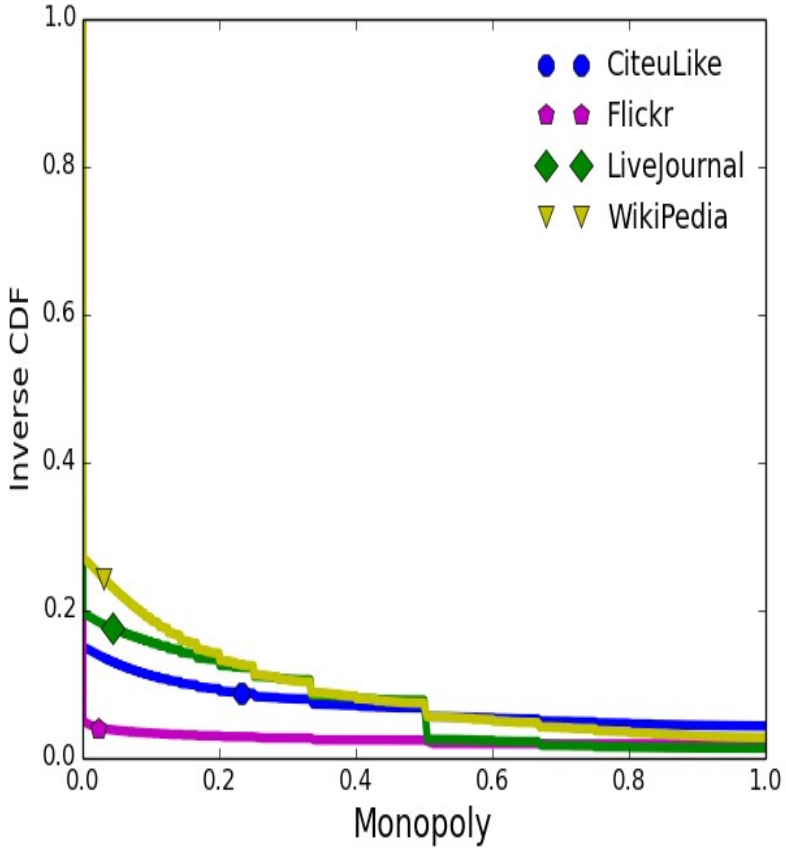
Strong correlation for CC
RD does not depend on the degree

Results

- Distribution of dispersion and monopoly



Distribution around high values

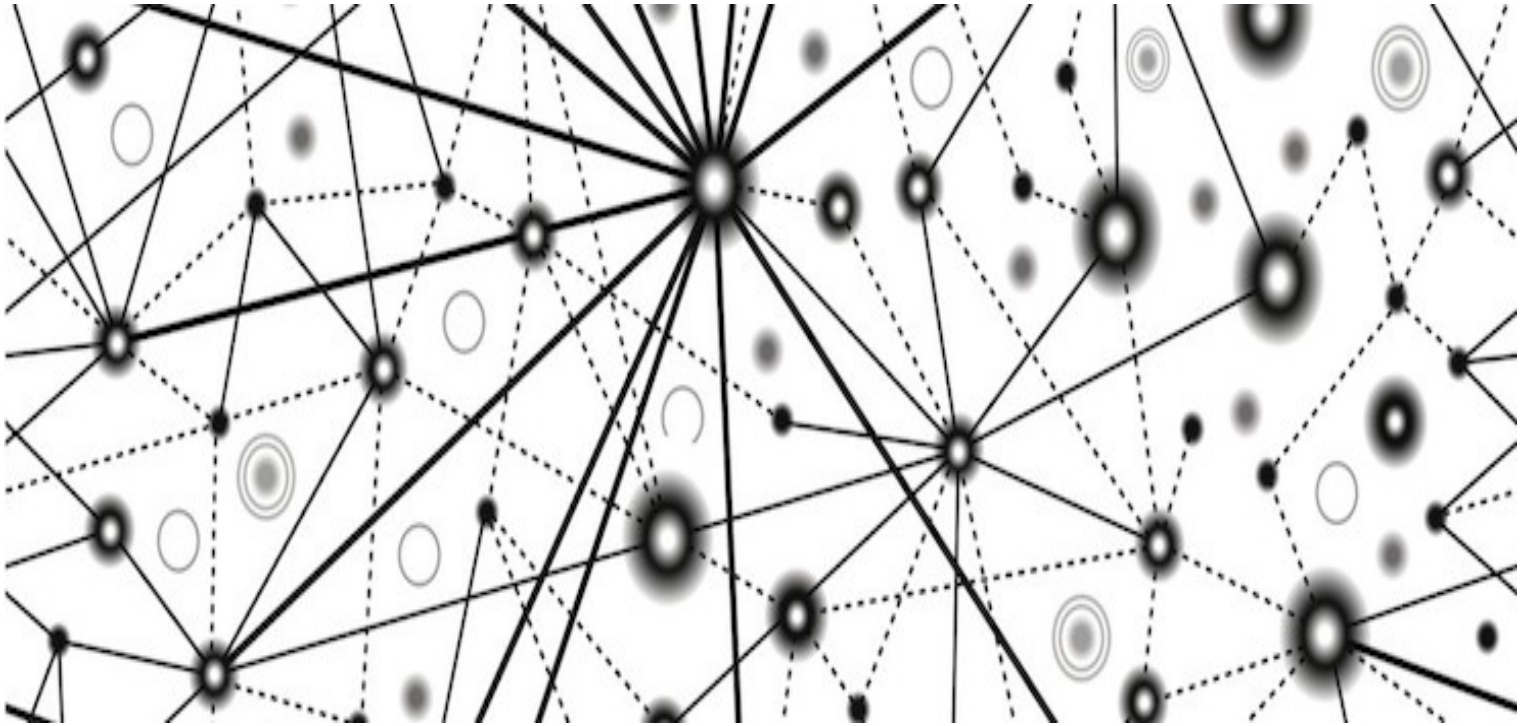


Very low monopolies

Conclusion and future work

- Conclusion
  - Online social networks present particularities
  - The overlap is not trivial
- Future work
  - Analysis of detected communities
  - Detection of communities in bipartite graph

Thanks !

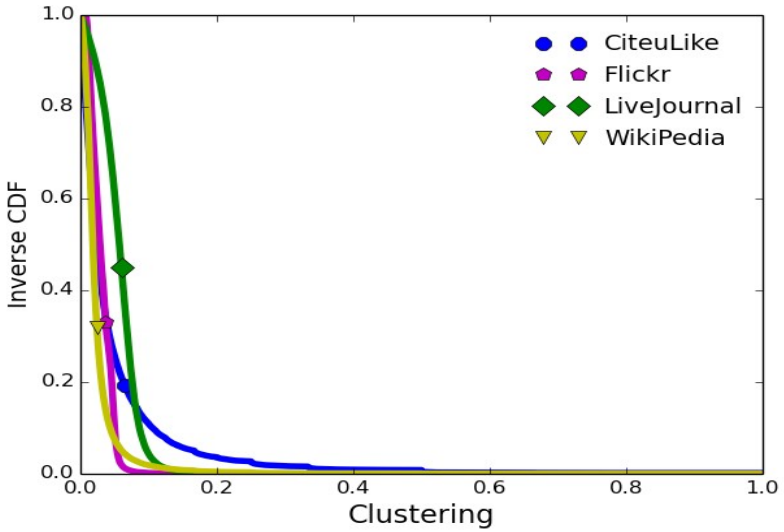


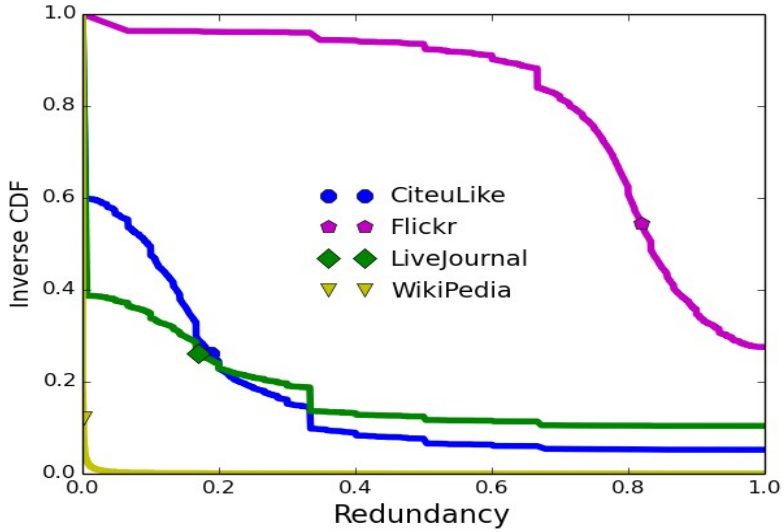
Results : random model

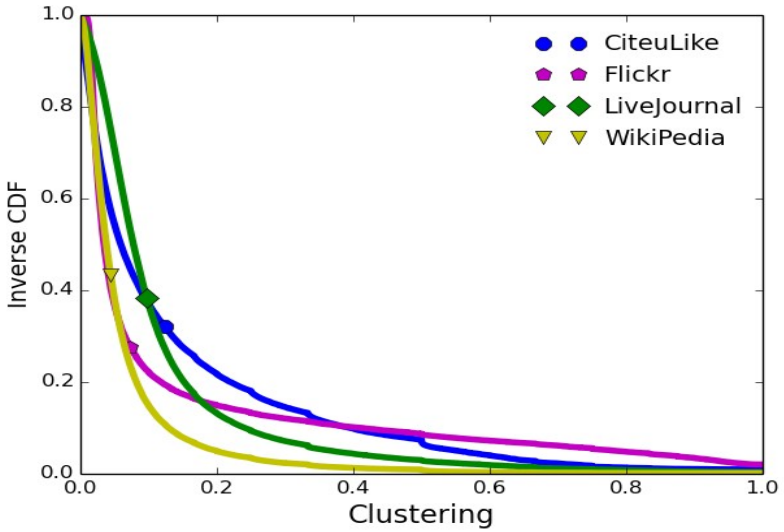
- Configuration Model method
- Goal : compare with real data
  - Caracterize specific properties
  - Denote differences
  - Understand how the metrics react

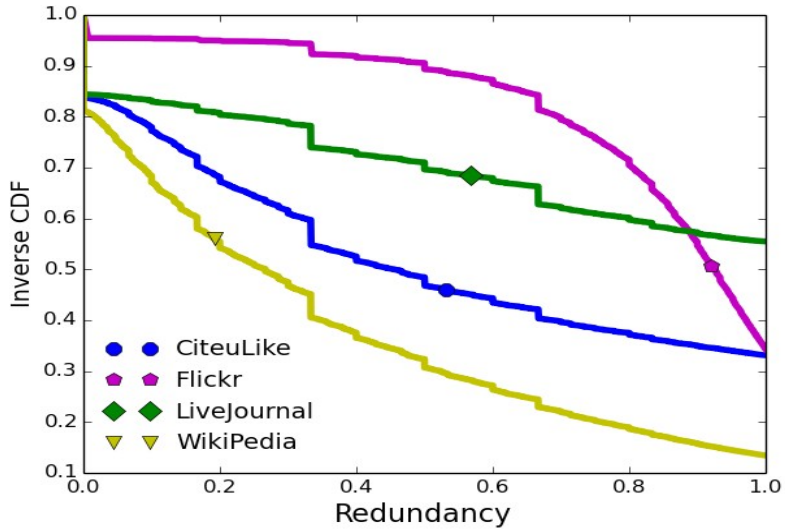
Results : random model

- Distribution of CC and RD



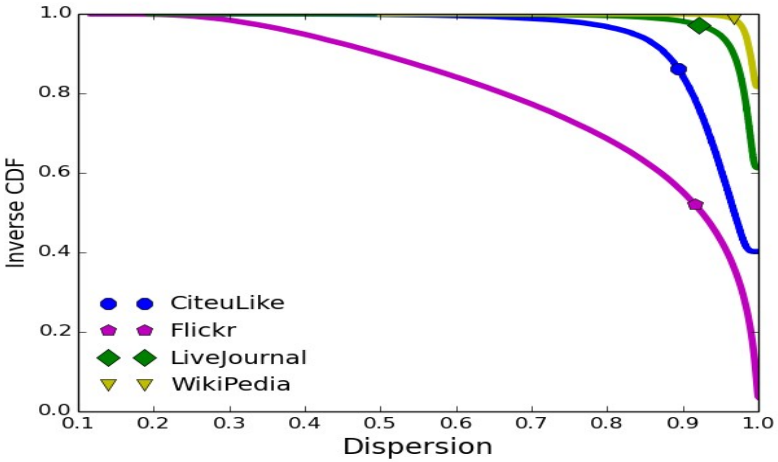


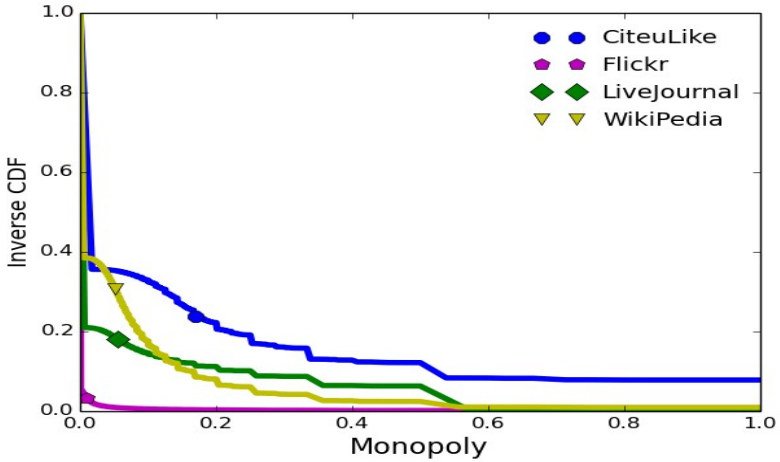


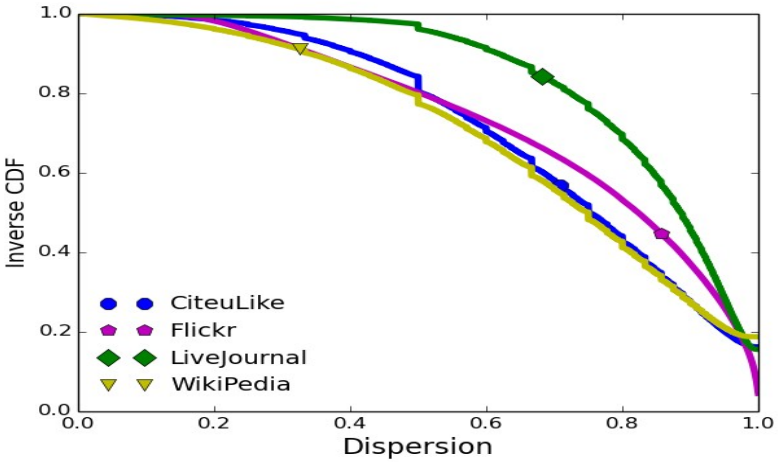


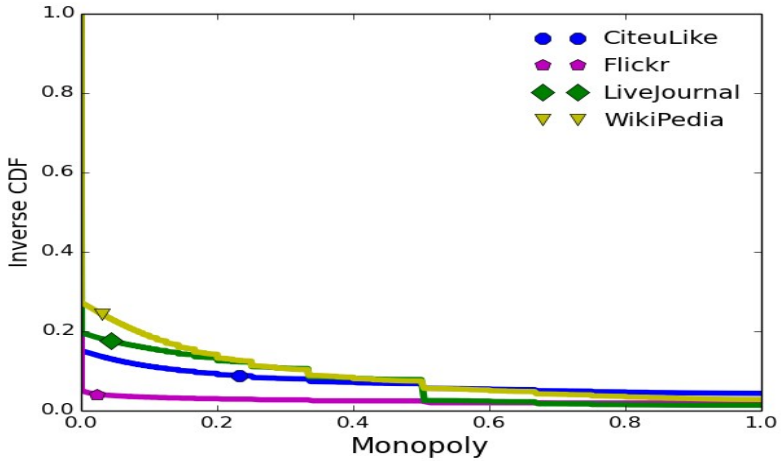
Results : random model

- Distribution of dispersion and monopoly









Results : random model

- Correlation between degree/CC and degree/RC