On the Permanence of Vertices in Network Communities

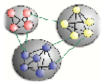
Tanmoy Chakraborty, Niloy Ganguly, Animesh Mukherjee
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur, India – 721302
http:// cnerg.org

Sriram Srinivasan, Sanjukta Bhowmick
Department of Computer Science
University of Nebraska, Omaha, Nebraska - 68182
http://cs.unomaha.edu/~bhowmick

http://cnerg.org/permanence

Introduction

- **Community:** Group of nodes within which the connection is dense, but between which the connection is relatively sparse



- Community structure indicates **structural or functional similarities** between nodes in a network
- **Identifying** and **analyzing** community structure are two fundamental research agenda since last 10 years

Background

❑ Community Finding Algorithms:
- **Modularity-based:** FastGreedy [Newman, 14], Louvain [Blondel et al, 08] and CNM [Clause et al, 04]
- **Random walk-based:** WalkTrap [Pons & Latapy, 06]
- **Compression-based:** InfoMod and InfoMap [Rosvall & Bergstrom, 07]

❑ Community Scoring Functions:
- Modularity [Newman, 06]
- Conductance [Leskovec et al., 09]
- Cut-ratio [Leskovec et al., 10]

Limitations

❑ Existing algorithms are prone to
- arbitrary network noise
- vertex ordering [Chakraborty et al, 13]
- initial seed node selection

❑ Other limitations [Good et al, 10]
- Resolution limit
- Degeneracy of solution
- Asymptotic growth

❑ No one measures the **degree of belongingness** of a vertex in its own community

❑ **Q:** Is a network **eligible** for community analysis?

Aims

❑ Defining suitable **community scoring metric** that
- minimizes existing limitations
- is sensitive to network perturbation
- qualifies for a standard community goodness measurement metric

❑ **By-product:** developing an **optimization algorithm** for detecting non-overlapping communities

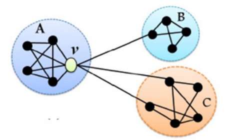
Proposed metric

Assumptions:
- a vertex should have **more internal connections** than the **maximum** connections to **any one** of the neighbouring community
- the **internal neighbours** should be **highly connected** among each other

Permanence

$$Perm(v) = \frac{I(v)}{D(v)} \times \frac{1}{E_{max}(v)} - (1 - C_{in}(v))$$

$I(v)$ = Internal neighbors of v
$D(v)$ = Degree of v
$E_{max}(v)$ = Max connections to an external community of v
$C_{in}(v)$ = Clustering co-efficient of internal neighbors of v



$I(v) = 4$, $D(v) = 7$, $E_{max}(v) = 2$, $C_{in}(v) = 5/6$
$Perm(v) = 0.12$

Test Suite of Networks

❑ Synthetic Networks:
LFR benchmark networks with given community structure [Lancichinetti & Fortunato, 2009]

❑ Real-world Networks
- **Football:** Nodes: teams, Edges: matches, Communities: team-conferences [Girvan & Newman, 02]
- **Railway:** Nodes: stations, Edges: railway connections, Communities: states/provinces [Ghosh et al., 11]
- **Coauthorship:** Nodes: authors, Edges: coauthorship, Communities: research area [Chakraborty et al, 13]

Properties	Football	Railway	Coauthorship
# nodes	115	301	103677
# edges	613	1224	352183
# communities	12	21	24
Smallest size community	5	1	34
Largest size community	13	46	14404

Validation Metrics to compare with ground-truth communities

- Normalized Mutual Information (NMI)
- Adjusted Rand Index (ARI)
- Purity (PU)
- Weighted NMI (W-NMI)
- Weighted ARI (W-ARI)
- Weighted PU (W-PU)

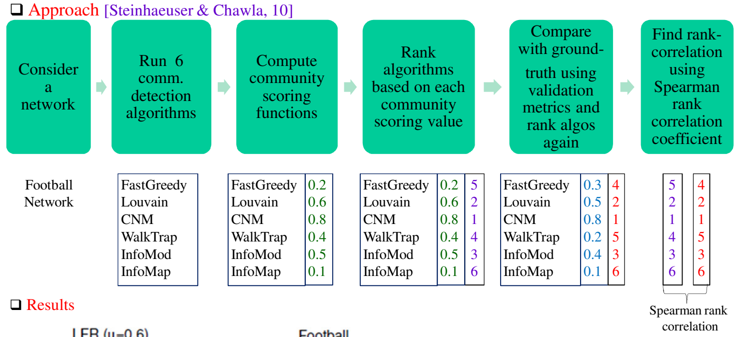
[Manning et al, 09] [Labatut, 13]

Discussions

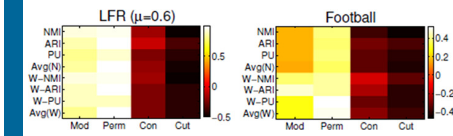
- Value of permanence correlates to the community structure
- Able to detect small-size communities
- Minimizes the resolution limit, asymptotic growth and degeneracy problems theoretically
- Indicates the eligibility of the network for community detection

Permanence: A Community Scoring Function

❑ Approach [Steinhaeuser & Chawla, 10]



❑ Results



Heat maps depicting pair-wise Spearman rank correlation

Networks	Modularity	Permanence	Conductance	Cut
LFR(μ=0.1)	**0.88**	**0.88**	**0.88**	0.02
LFR(μ=0.3)	0.61	**0.74**	0.72	0.28
LFR(μ=0.6)	0.87	**0.96**	-0.18	-0.44
Football	0.25	**0.43**	-0.29	-0.41
Railway	0.43	**0.46**	0.08	-0.48
Coauthorship	**0.92**	**0.92**	0.76	0.86

Performance of the community scoring functions averaged over all the validation measures

Maximizing Permanence for Community Detection

- Follow similar strategy used in Louvain algorithm (a greedy modularity maximization) [Blondel et al., 07]
- Selecting seed nodes helps converge the process faster
- We only consider those communities having size >=3
- Communities having size <3 remain as singleton

❑ **Results:** Differences of our algorithm with the other algorithms averaged over all validation measures

Algorithms	LFR (μ=0.1)	LFR (μ=0.3)	LFR (μ=0.6)	Football	Railway	Coauthorship
Louvain	0.02	0.00	-0.75	0.02	0.14	0.00
FastGreedy	0.00	0.87	0.02	0.01	0.37	0.14
CNM	0.06	0.40	??? -0.13	0.30	0.00	0.05
WalkTrap	0.00	0.00	-0.50	0.02	0.02	0.01
Infomod	0.11	0.08	-0.20	0.19	0.04	0.00
Infomap	0.00	0.00	-0.72	0.02	-0.02	0.03

Acknowledgements

Google India Fellowship, College of IS&T, Graca grant for UNO-SPR, RISC Fellowship, MSR and Yahoo travel grants