On the permanence of vertices in network communities

Tanmoy Chakraborty[1,a], Sriram Srinivasan[2,b], Niloy Ganguly[3,a], Animesh Mukherjee[4,a], Sanjukta Bhowmick[5,b]

[a]*Dept. of Computer Science & Engg., Indian Institute of Technology, Kharagpur, India – 721302;*

[b]*Dept. of Computer Science, University of Nebraska, Omaha, Nebraska 68182*

{[1]*its_tanmoy*,[3]*niloy*,[4]*animeshm*}*@cse.iitkgp.ernet.in*

{[2]*sriramsrinivas*,[5]*sbhowmick*}*@unomaha.edu*

Appendix

1. Proofs of the Lemmas

Here we present the detailed proofs of the Lemmas discussed in the main text. For the explanations, we again show the toy example in Figure 1.

We consider two communities A and B to be connected via a vertex v. The vertex v has connections to α nodes in community A and to β nodes in community B, and these nodes form the set N_A and N_B respectively. The average internal degree of a vertex, $a \in N_A$ ($b \in N_B$), before v is added is I_A (I_B). Similarly, the average internal clustering coefficient of a vertex, $a \in N_A$ ($b \in N_B$), before v is added is C_A (C_B). We assume the values of C_A and C_B to be at least 0.5. Communities A and B have no other connections except those through v. We also assume that $\alpha \geq \beta$

When v is added to communities A (B) then the average internal clustering coefficient of v becomes C^v_A (C^v_B), and the average clustering coefficient of the nodes in $N_A(N_B)$ become C^α (C^β). We consider two extreme values of $C^\alpha(C^\beta)$. One case is when the nodes in the community are tightly connected and adding v does not significantly change the internal clustering coefficient. In this case, we assume $C^\alpha = C_A$ and $C^\beta = C_B$. The other case is when the nodes in the community are not as tightly connected. In this case, adding v decreases the average internal clustering coefficient.

Let the number of internal connections of nodes in N_A, before v is added, be f_x. Therefore $C_A = \frac{fx}{I_A(I_A-1)}$. In the second case when the communities are sparse, once v is added we assume that no new distance two connections are formed, but the internal degree increases by one. Therefore $C^\alpha = \frac{fx}{(I_A+1)I_A} = \frac{C_A I_A(I_A-1)}{(I_A+1)I_A} = \frac{C_A(I_A-1)}{(I_A+1)}$. Similarly $C^\beta = \frac{C_B(I_B-1)}{(I_B+1)}$.

The combination of communities A, B and the vertex v can have four cases as follows:

- **Case 1.** v joins with community A only. We denote this configuration as

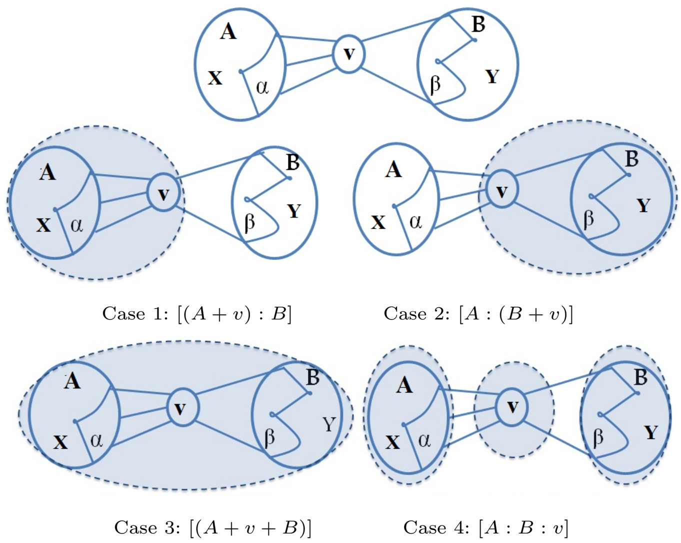

Fig. 1. (Color online) Toy example demonstrating four Lemmas.

$[(A+v):B]$, and its total permanence as $P_{(A+v):B}$. We assume that the combined permanence of all nodes $x \notin (N_A \cup N_B \cup v)$ as P_x. This value will not be affected due to the re-assignments. Therefore, the total permanence is the sum of the following factors: P_x, $[\alpha C^{\alpha}]$ (for the nodes in N_A connected to v), $[\frac{\alpha}{(\alpha+\beta)\beta} - (1-C_A^v)]$ (for vertex v) and $[\beta(\frac{I_B}{I_B+1} - (1-C^{\beta}))]$ (for the nodes in N_B).

$$P_{(A+v):B} = P_x + \alpha C^{\alpha} + \frac{\alpha}{(\alpha+\beta)\beta} - (1-C_A^v) + \beta(\frac{I_B}{I_B+1} - (1-C_B))$$

- **Case 2.** v joins with community B only. We denote this configuration as $[(A:(v+B)]$, and its total permanence as $P_{A:(v+B)}$. The values of this total permanence is the sum of the following factors: P_x, $[\alpha(\frac{I_A}{I_A+1} - (1-C^{\alpha}))]$ (for the nodes in N_A), $[\frac{\beta}{(\alpha+\beta)\alpha} - (1-C_B^v)]$ (for vertex v) and $[\beta C^{\beta}]$ (for the nodes in N_B connected to v).

$$P_{A:(v+B)} = P_x + \alpha(\frac{I_A}{I_A+1} - (1-C_A)) + \frac{\beta}{(\alpha+\beta)\alpha} - (1-C_B^v) + \beta C^{\beta}$$

- **Case 3.** A, B and v merge together. We denote this configuration as $[(A+v+B)]$, and its total permanence as $P_{(A+v+B)}$. The values of this total permanence is the sum of the following factors: P_x, $[\alpha C^{\alpha}]$ (for the nodes in N_A), $[\frac{\alpha(\alpha-1)C_A^v+\beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)}]$ (for vertex v) and $[\beta C^{\beta}]$ (for the nodes in N_B connected to v).

$$P_{(A+v+B)} = P_x + \alpha C^{\alpha} + \frac{\alpha(\alpha-1)C_A^v+\beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} + \beta C^{\beta}$$

- **Case 4.** A, B and v remain as separate communities. We denote this configuration as $[(A : v : B)]$, and its total permanence as $P_{(A:v:B)}$. The values of this total permanence is the sum of the following factors: P_x, $[\alpha(\frac{I_A}{I_A+1} - (1 - C^{\alpha}))]$ (for the nodes in N_A), 0 (for vertex v) and $[\beta(\frac{I_B}{I_B+1} - (1 - C^{\beta}))]$ (for the nodes in N_B).

  $P_{(A:v:B)} = P_x + \alpha(\frac{I_A}{I_A+1} - (1 - C_A)) + \beta(\frac{I_B}{I_B+1} - (1 - C_B))$

Lemma 4.1 *Given $C^{\alpha} = C_A$ and $C^{\beta} = C_B$, let $Z = \frac{\alpha-\beta}{\alpha\beta} + (C_A^v - C_B^v) + \left(\frac{\alpha}{I_A+1} - \frac{\beta}{I_B+1}\right)$. The assignment $[(A+v) : B]$ will have a higher permanence than $[A : (v+B)]$, if $Z > 0$ and a lower permanence if $Z < 0$.*

Proof. Here we are comparing between Case 1 and Case 2. The difference in total permanence between these two assignments by assuming $C^{\alpha} = C_A$ and $C^{\beta} = C_B$ is:

$$
\begin{aligned}
P_{(A+v):B} - P_{A:(v+B)} &= \frac{\alpha}{(\alpha+\beta)\beta} + C_A^v + \beta(\frac{I_B}{I_B+1} - 1) \\
&- (\alpha(\frac{I_A}{I_A+1} - 1) + \frac{\beta}{(\alpha+\beta)\alpha} + C_B^v) \\
&= \frac{\alpha}{(\alpha+\beta)\beta} - \frac{\beta}{(\alpha+\beta)\alpha} \\
&+ (C_A^v - C_B^v) + (-\beta\frac{1}{I_B+1} - \alpha\frac{-1}{I_A+1}) \\
&= \frac{\alpha-\beta}{\alpha\beta} + (C_A^v - C_B^v) + (\frac{\alpha}{I_A+1} - \frac{\beta}{I_B+1})
\end{aligned}
$$

If this difference is greater than zero then $[(A+v) : B]$ will have a higher permanence. If the difference is less than zero then $[A : (v+B)]$ will have higher permanence.

Lemma 4.2 Merging the communities A, B and v, gives higher permanence than joining v to community A if $C^{\beta} = C_B$, and $\frac{\gamma}{(\gamma+1)\beta} + \frac{C_A^v(2\gamma+1)-C_B^v}{(\gamma+1)^2} - \frac{\beta}{I_B+1} < 1$; where $\gamma = \alpha/\beta$, and also if $C^{\beta} = C_B\frac{I_B-1}{I_B+1}$, and $\frac{\gamma}{(\gamma+1)\beta} + \frac{C_A^v(2\gamma+1)-C_B^v}{(\gamma+1)^2} + \frac{\beta(2C_B-1)}{I_B+1} < 1$.

Proof. We are comparing Case 1 and Case 3 and in this case $C^{\beta} = C_B\frac{I_B-1}{I_B+1}$.

The difference in total permanence is:

$$
\begin{aligned}
P_{(A+v):B} - P_{(A+v+B)} &= \frac{\alpha}{(\alpha+\beta)\beta} - 1 + C_A^v + \beta(\frac{I_B}{I_B+1} - 1 + C_B) \\
&- (\frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} + \beta C^\beta) \\
&= \frac{\alpha}{(\alpha+\beta)\beta} - 1 + C_A^v - \frac{\beta}{I_B+1} + \beta(C_B - C^\beta) \\
&- \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} \\
&\text{Substituting } \gamma = \alpha/\beta \text{ and } C^\beta = C_B \tfrac{I_B-1}{I_B+1} \\
&= \frac{\gamma}{(\gamma+1)\beta} - 1 + C_A^v - \frac{\beta}{I_B+1} + \beta C_B \frac{2}{I_B+1} \\
&- \frac{\gamma(\gamma-1/\beta)C_A^v + (1-1/\beta)C_B^v}{(\gamma+1)(\gamma+1-1/\beta)}
\end{aligned}
$$

The value of $1/\beta$ will become lower as β increases. We therefore ignore its effect. The equation then becomes:

$$
\begin{aligned}
P_{(A+v):B} - P_{(A+v+B)} &= \frac{\gamma}{(\gamma+1)\beta} - 1 + C_A^v - \frac{\beta}{I_B+1} + \beta C_B \frac{2}{I_B+1} \\
&- \frac{\gamma^2 C_A^v}{(\gamma+1)(\gamma+1)} - \frac{C_B^v}{(\gamma+1)(\gamma+1)} \\
&= \frac{\gamma}{(\gamma+1)\beta} - 1 + \frac{C_A^v(2\gamma+1) - C_B^v}{(\gamma+1)^2} + \beta \frac{2C_B - 1}{I_B+1}
\end{aligned}
$$

If this difference is less than 0 then higher permanence is obtained by merging. Therefore, the condition to merge A, B and v altogether rather than v joining with A is:
$\frac{\gamma}{(\gamma+1)\beta} + \frac{C_A^v(2\gamma+1)-C_B^v}{(\gamma+1)^2} + \beta\frac{2C_B-1}{I_B+1} < 1$

If we consider $C^\beta = C_B$, then

$$
P_{(A+v):B} - P_{(A+v+B)} = \frac{\gamma}{(\gamma+1)\beta} - 1 + \frac{C_A^v(2\gamma+1) - C_B^v}{(\gamma+1)^2} - \frac{\beta}{I_B+1}
$$

In this case, the condition to merge is:

$$
\frac{\gamma}{(\gamma+1)\beta} + \frac{C_A^v(2\gamma+1) - C_B^v}{(\gamma+1)^2} - \frac{\beta}{I_B+1} < 1
$$

`Corollary 4.5` *If $\beta = 1$, $C^\beta = C_B \frac{I_B-1}{I_B+1}$, $C_A^v > 1/2$ then v will join community A rather than the three pieces merging.*

`Corollary 4.6` *If $C_B \approx 1$, $C^\beta = C_B \frac{I_B-1}{I_B+1}$, $\beta \geq I_B + 1$ and $C_A^v \geq C_B^v/3$ then v will join community A rather than the three pieces merging.*

Proof of Corollary 4.5: If $\beta = 1$, then

$$
\begin{aligned}
P_{(A+v):B} - P_{(A+v+B)} &= \frac{\gamma}{(\gamma+1)\beta} - 1 + C_A^v - \frac{\beta}{I_B+1} \\
&+ \beta C_B \frac{2}{I_B+1} - \frac{(\gamma(\gamma-1/\beta)C_A^v + (1-1/\beta)C_B^v}{(\gamma+1)(\gamma+1-1/\beta)} \\
&= \frac{\gamma}{(\gamma+1)} - 1 + C_A^v + \frac{(2C_B-1)}{I_B+1} \\
&- \frac{\gamma(\gamma-1)C_A^v}{(\gamma+1)(\gamma)} \\
&= \frac{-1}{(\gamma+1)} + C_A^v \frac{2\gamma}{\gamma+1)\gamma} + \frac{(2C_B-1)}{I_B+1} \\
&= \frac{2C_A^v - 1}{(\gamma+1)} + \frac{(2C_B-1)}{I_B+1}
\end{aligned}
$$

This value will be positive so long as $C_A^v > 1/2$. In this case, joining v to community A is favored.

Proof of Corollary 4.6: If $C_B \approx 1$, $\beta \geq I_B + 1$ and $C_A^v \geq C_B^v$, then

$$
\begin{aligned}
P_{(A+v):B} - P_{(A+v+B)} &= \frac{\gamma}{(\gamma+1)\beta} - 1 + C_A^v - \frac{\beta}{I_B+1} \\
&+ \beta C_B \frac{2}{I_B+1} - \frac{\gamma(\gamma-1/\beta)C_A^v + (1-1/\beta)C_B^v}{(\gamma+1)(\gamma+1-1/\beta)} \\
&\text{Ignore } 1/\beta\text{, because } \beta \text{ is high} \\
&= \frac{\gamma}{(\gamma+1)\beta} - 1 + C_A^v + \beta \frac{2C_B-1}{I_B+1} \\
&- \frac{\gamma(\gamma-1/\beta)C_A^v}{(\gamma+1)(\gamma+1)} - \frac{C_B^v}{(\gamma+1)(\gamma+1)} \\
&= \frac{\gamma}{(\gamma+1)\beta} + \frac{C_A^v(2\gamma+1) - C_B^v}{(\gamma+1)(\gamma+1)} \\
&+ (\frac{\beta}{I_B+1} - 1)
\end{aligned}
$$

The first term is positive. Since the smallest value of $\gamma = 1$, and $C_A^v \geq C_B^v$, the second term is positive, and since $\beta \geq I_B + 1$, the third term is also positive. Therefore v will join community A rather than merging all the components.

Lemma 4.3 *If $C^\alpha = C_A$ and $C^\beta = C_B$ then the communities will merge (i.e., $[(A+v+B)]$), rather than remain separate (i.e., $[A : B : C]$). If $C^\alpha = C_A \frac{(I_A-1)}{(I_A+1)}$ $C^\beta = C_B \frac{(I_B-1)}{(I_B+1)}$ and then the communities will merge if:*

$\frac{\gamma^2 C_A^v + C_B^v}{(\gamma+1)^2} > \alpha \frac{(2C_A - 1)}{I_A+1} + \beta \frac{(2C_B-1)}{I_B+1}$.

Proof: We are comparing Case 3 and Case 4, and the case $C^\beta = C_B \frac{I_B-1}{I_B+1}$. The difference in total permanence is:

$$
\begin{aligned}
P_{(A+v+B)} - P_{(A:v:B)} &= \alpha C^\alpha + \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} + \beta C^\beta \\
&- (\alpha(\frac{I_A}{I_A+1} - (1 - C_A)) \\
&+ \beta(\frac{I_B}{I_B+1} - (1 - C_B))) \\
&= -\alpha(C_A - C^\alpha) - \beta(C_B - C^\beta) \\
&+ \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} \\
&+ \frac{\alpha}{I_A+1} + \frac{\beta}{I_B+1} \\
&= \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} \\
&+ \frac{\alpha}{I_A+1} + \frac{\beta}{I_B+1} \\
&- (\alpha\frac{2C_A}{I_A+1} + \beta\frac{2C_B}{I_B+1}) \\
&\text{Substituting } \gamma = \alpha/\beta \\
&= \frac{\gamma(\gamma - 1/\beta)C_A^v + (1 - 1/\beta)C_B^v}{(\gamma+1)(\gamma+1-1/\beta)} \\
&- (\alpha\frac{(2C_A-1)}{I_A+1} + \beta\frac{(2C_B-1)}{I_B+1})
\end{aligned}
$$

The value of $1/\beta$ will become lower as β increases. We therefore ignore its effect. The equation then becomes;

$$
\begin{aligned}
P_{(A+v+B)} - P_{(A:v:B)} = \frac{\gamma^2 C_A^v + C_B^v}{(\gamma+1)^2} \\
- (\alpha\frac{(2C_A-1)}{I_A+1} + \beta\frac{(2C_B-1)}{I_B+1})
\end{aligned}
$$

If $C^\alpha = C_A$ and $C^\beta = C_B$, then

$$\begin{aligned} P_{(A+v+B)} - P_{(A:v:B)} &= \alpha C_A + \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} + \beta C_B \\ &- (\alpha(\frac{I_A}{I_A+1} - (1-C_A)) \\ &+ \beta(\frac{I_B}{I_B+1} - (1-C_B))) \\ &= \frac{\alpha(\alpha-1)C_A^v + \beta(\beta-1)C_B^v}{(\alpha+\beta)(\alpha+\beta-1)} \\ &+ \frac{\alpha}{I_A+1} + \frac{\beta}{I_B+1} \end{aligned}$$

This value is always positive so the communities will merge.

Lemma 4.4 *If $C^\alpha = C_A$ and $C^\beta = C_B$ then the communities will remain separate (i.e., $[A : v : B]$) rather than v joining with community A (i.e., $[(A + v) : B]$), if $\alpha(\frac{1}{I_A+1} + \frac{1}{(\alpha+\beta)\beta}) < (1 - C_A^v)$.*
Otherwise, if $C^\alpha = C_A \frac{(I_A-1)}{(I_A+1)}$; $C^\beta = C_B \frac{(I_B-1)}{(I_B+1)}$ and then the communities will remain separate if $\alpha(\frac{2C_A-1}{I_A+1}) + (1 - C_A^v) \geq \frac{\alpha}{(\alpha+\beta)\beta}$

Proof: We are comparing Case 1 and Case 4 for the case $C^\alpha = C_A \frac{(I_A-1)}{(I_A+1)}$; $C^\beta = C_B \frac{(I_B-1)}{(I_B+1)}$. The difference in total permanence is:

$$\begin{aligned} P_{(A+v):B} - P_{(A:v:B)} &= \alpha C^\alpha + \frac{\alpha}{(\alpha+\beta)\beta} - (1 - C_A^v) \\ &- (\alpha(\frac{I_A}{I_A+1} - (1-C_A)) \\ &= \alpha(C^\alpha - C_A + \frac{1}{I_A+1}) + \frac{\alpha}{(\alpha+\beta)\beta} \\ &+ (C_A^v - 1) \\ &= \alpha(\frac{1-2C_A}{I_A+1}) + \frac{\alpha}{(\alpha+\beta)\beta} \\ &+ (C_A^v - 1) \end{aligned}$$

This value will be negative (favor merge) if: $\alpha(\frac{2C_A-1}{I_A+1} + (1 - C_A^v)) > \frac{\alpha}{(\alpha+\beta)\beta}$

If we consider the case $C^\alpha = C_A$ and $C^\beta = C_B$, then

$$P_{(A+v):B} - P_{(A:v:B)} = \alpha C^{\alpha} + \frac{\alpha}{(\alpha+\beta)\beta} - (1 - C_A^v)$$
$$- (\alpha(\frac{I_A}{I_A+1} - (1 - C_A))$$
$$= \frac{\alpha}{I_A+1} + \frac{\alpha}{(\alpha+\beta)\beta}$$
$$+ (C_A^v - 1)$$
$$= \alpha(\frac{1}{I_A+1} + \frac{1}{(\alpha+\beta)\beta})$$
$$+ (C_A^v - 1)$$

This value will be negative (favor merge) if: $\alpha(\frac{1}{I_A+1}) + \frac{1}{(\alpha+\beta)\beta}) < (1 - C_A^v)$

`Corollary 4.7` *If* $\alpha = \beta$, $C^{\beta} = C_B \frac{I_B-1}{I_B+1}$, $C_A^v = C_B^v$ *then communities A, B and v will merge, rather than v joining community A, if* $\frac{1}{2\beta} + \frac{C_A^v}{2} + \beta\frac{2C_B-1}{I_B+1} < 1$.

`Corollary 4.8` *If* $\alpha = \beta$, $C^{\beta} = C_B \frac{I_B-1}{I_B+1}$, *then communities A, B and v will remain separate rather than v joining community A, if* $\alpha(\frac{2C_A-1}{I_A+1}) + (1 - C_A^v) \geq \frac{1}{2\alpha}$. *If* $\alpha = \beta = 1$, *then* $C_A^v = 0$ *the communities will remain always separated.*

Proof of Corollary 4.7: If $\alpha = \beta$, $C_A^v = C_B^v$, $C^{\beta} = C_B \frac{I_B-1}{I_B+1}$, then comparing Case 1 and Case 3 we get

$$P_{(A+v+B)} - P_{(A:v:B)} = \frac{\gamma}{(\gamma+1)\beta} - 1 + \frac{C_A^v(2\gamma+1) - C_B^v}{(\gamma+1)^2} + \beta\frac{2C_B-1}{I_B+1}$$
$$= \frac{1}{2\beta} - 1 + \frac{3C_A^v - C_A^v}{4} + \beta\frac{2C_B-1}{I_B+1}$$
$$= \frac{1}{2\beta} - 1 + \frac{2C_A^v}{4} + \beta\frac{2C_B-1}{I_B+1}$$

This values is negative(favors merging) if $\frac{1}{2\beta} + \frac{C_A^v}{2} + \beta\frac{2C_B-1}{I_B+1} < 1$

Proof of Corollary 4.8: If $\alpha = \beta$, $C^{\beta} = C_B \frac{I_B-1}{I_B+1}$, $C_A^v = C_B^v$ then;

$$P_{(A+v):B} - P_{(A:v:B)} = \alpha(\frac{1-2C_A}{I_A+1}) + \frac{\alpha}{(\alpha+\beta)\beta}$$
$$+ (C_A^v - 1)$$
$$= \alpha(\frac{1-2C_A}{I_A+1}) + \frac{1}{2\alpha}$$
$$+ (C_A^v - 1)$$

This value will be negative (favor merge) if: $\alpha(\frac{2C_A-1}{I_A+1})+(1-C_A^v))>\frac{1}{2\alpha}$

If $\alpha=1$, then $C_A^v=0$. Then the condition is: $\frac{2C_A-1}{I_A+1}+1>\frac{1}{2}$, Which is always true.

2. Other Ranking Schemes for Evaluating Community Scoring Functions

Besides dense ranking that is shown in the main text (section 6), we use two other ranking schemes, namely standard competition ranking and fractional ranking to measure the performance of permanence over other community scoring functions for evaluating the goodness of the detected communities. The results are shown in the following two subsections.

2.1. *Standard competition ranking*

In standard competition ranking[a], values that are equal receive the same ranking number, and then a gap is left in the ranking numbers. The number of ranking numbers that are left out in this gap is one less than the number of items that are equal. For instance, if the values are [8 7 7 2 1], the corresponding ranks are [1 2 2 4 5]. For football network, the standard ranks of the algorithms based on the community scoring functions and validation measures are shown in Table 1. Similarly, we measure such ranks for all the networks and find out the Spearman's rank correlation [1] between all pairs of scoring functions and validation measures. The heat maps in Figure 2 depict the rank correlations for six different networks namely, LFR ($\mu=0.1$), LFR ($\mu=0.3$), LFR ($\mu=0.6$), football, railway and coauthorship networks. This is best captured in Table 2 which presents the average values of these community scoring functions computed over all the validation measures for each of the networks. We can observe that here also in all the cases, the correlation is highest for permanence measure.

2.2. *Fractional ranking*

In fractional ranking[b], values that are equal receive the same ranking number, which is the *mean* of what they would have under ordinal rankings[c]. For instance, suppose we have the data set [1 1 2 3 3 4 5 5 5]. There are 5 different numbers, so there would be five different ranks. If first two numbers, i.e., 1 and 1 were actually different numbers, they would occupy ranks 1 and 2. Since they are the same number, we find their ranks by finding the average as follows: [(rank) 1 + (rank) 2] / 2 numbers total = 1.5 (average rank). The next number in the data set, 2, is thus assigned

[a]http://en.wikipedia.org/wiki/Ranking
[b]http://en.wikipedia.org/wiki/Ranking
[c]in ordinal ranking, all items receive distinct ordinal numbers, including items that compare equal.

Table 1. For Football network, the values of four community scoring functions on the output obtained from eight different algorithms and the scores of the validation measures with respect to the ground-truth communities. The ranks (using **standard ranking scheme**) of the algorithms based on individual measures are shown within parenthesis. The average ranks of all the normal (weighted) validation measures are shown in column 9 (column 13)

Algorithms	Mod	Perm	1-Con	1-Cut	NMI	ARI	PU	Avg (N)	W-NMI	W-ARI	W-PU	Avg (W)
Louvain	0.60(1)	0.36(1)	0.77 (5)	0.44(6)	0.93(1)	0.99(1)	0.89(2)	1.67	0.99(2)	0.93(2)	0.99(1)	1.67
FastGreedy	0.58(5)	0.25(5)	0.81(4)	0.59(4)	0.93(1)	0.99(1)	0.91(1)	1	1(1)	0.94(1)	0.99(1)	1
CNM	0.55(6)	0.20(6)	0.85(1)	0.86(1)	0.67(6)	0.75(6)	0.42(6)	6	0.55(6)	0.63(6)	0.71(6)	6
WalkTrap	0.60(1)	0.36(1)	0.82(2)	0.69(2)	0.90(3)	0.98(5)	0.84(3)	3.67	0.98(3)	0.91(3)	0.99(1)	2.33
Infomod	0.60(1)	0.35(3)	0.82(2)	0.69(2)	0.89(4)	0.97(3)	0.82(4)	3.67	0.97(4)	0.89(4)	0.98(4)	4
Infomap	0.60(1)	0.35(3)	0.79(6)	0.51(5)	0.89(4)	0.97(3)	0.82(4)	3.67	0.97(4)	0.89(4)	0.98(4)	4

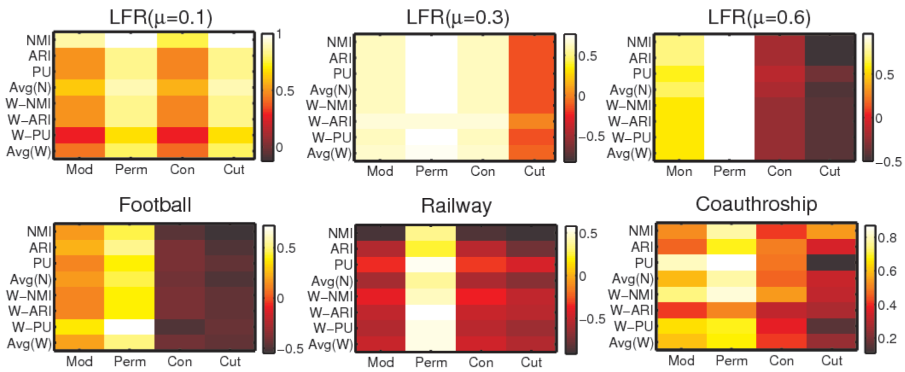

Fig. 2. Heat maps depicting the pairwise Spearman's rank correlation between four scoring functions with six validation measures for six different networks using **standard ranking scheme**. Avg(N) and Avg(W) are the averages of the ranks of three normal and three weighted validation measures as shown in Table 1.

Table 2. Performance of the community scoring functions averaged over all the validation measures for each network using **standard ranking scheme**. In each row, the highest correlation is highlighted in bold font.

Networks	Modularity	Permanence	Conductance	Cut
LFR(μ=0.1)	0.52	**0.84**	0.83	0.17
LFR(μ=0.3)	0.63	**0.74**	-0.11	-0.79
LFR(μ=0.6)	0.62	**0.95**	-0.19	-0.43
Football	0.18	**0.48**	-0.42	-0.47
Railway	-0.53	**0.44**	-0.51	-0.62
Coauthorship	0.6	**0.72**	0.43	0.29

the rank of 3 (the average takes up 1 and 2 in the first two 1's). The two 3's in the set would occupy ranks 3 and 4 if they were different numbers, so the average rank would be computed as follows: $(3 + 4)/2 = 3.5$. The value 4 would get the rank of 6 (because our average took into account rank 4 and 5 in the average). There are 3 5's in the data set. Their average rank is computed as $(7+8+9)/3 = 8$. Final ranks

Table 3. For Football network, the values of four community scoring functions on the output obtained from eight different algorithms and the scores of the validation measures with respect to the ground-truth communities. The ranks (using **fractional ranking scheme**) of the algorithms based on individual measures are shown within parenthesis. The average ranks of all the normal (weighted) validation measures are shown in column 9 (column 13)

Algorithms	Mod	Perm	1-Con	1-Cut	NMI	ARI	PU	Avg (N)	W-NMI	W-ARI	W-PU	Avg (W)
Louvain	0.60(2.5)	0.36(2.5)	0.77(6)	0.40(6)	0.93(1.5)	0.99(1.5)	0.89(2)	1.66	0.99(2)	0.93(2)	0.99(2)	2
FastGreedy	0.58(5)	0.25(5)	0.81(4)	0.59(4)	0.93(1.5)	0.99(1.5)	0.91(1)	1.33	1(1)	0.94(1)	0.99(2)	1.33
CNM	0.55(6)	0.20(6)	0.85(1)	0.86(1)	0.67(6)	0.75(6)	0.42(6)	6	0.55(6)	0.63(6)	0.71(6)	6
WalkTrap	0.60(2.5)	0.36(2.5)	0.82(2.5)	0.69(2.5)	0.90(3)	0.98(3)	0.84(3)	3	0.98(3)	0.91(3)	0.99(2)	2.66
Infomod	0.60(2.5)	0.35(3.5)	0.82(2.5)	0.69(2.5)	0.89(4.5)	0.97(4.5)	0.82(4.5)	4.5	0.97(4.5)	0.89(4.5)	0.98(4.5)	4.5
Infomap	0.60(2.5)	0.35(3.5)	0.79(5)	0.51(5)	0.89(4.5)	0.97(4.5)	0.82(4.5)	4.5	0.97(4.5)	0.89(4.5)	0.98(4.5)	4.5

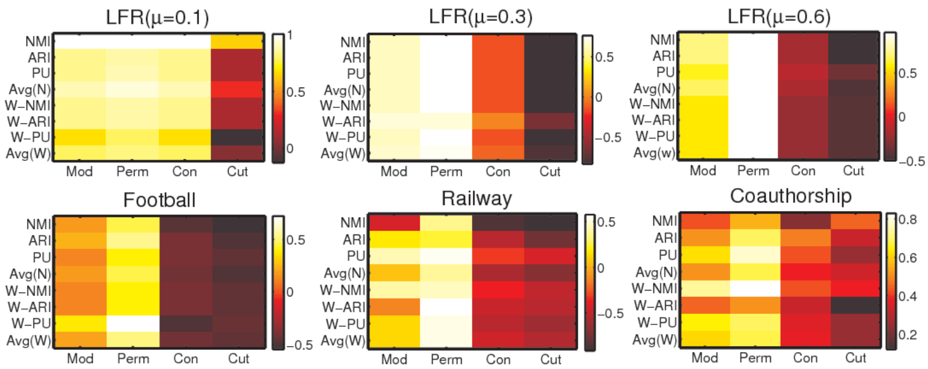

Fig. 3. Heat maps depicting the pairwise Spearman's rank correlation between four scoring functions with six validation measures for six different networks using **fractional ranking scheme**. Avg(N) and Avg(W) are the averages of the ranks of three normal and three weighted validation measures as shown in Table 3.

Table 4. Performance of the community scoring functions averaged over all the validation measures for each network **fractional ranking scheme**. In each row, the highest correlation is highlighted in bold font.

Networks	Modularity	Permanence	Conductance	Cut
LFR(μ=0.1)	0.83	**0.87**	0.83	0.16
LFR(μ=0.3)	0.63	**0.74**	-0.11	-0.79
LFR(μ=0.6)	0.62	**0.95**	-0.19	-0.43
Football	0.18	**0.48**	-0.42	-0.47
Railway	0.09	**0.44**	-0.51	-0.62
Coauthorship	0.56	**0.67**	0.37	0.29

would be [1.5 1.5 3 4.5 4.5 6 8 8 8].

For football network, the fractional ranks of the algorithms based on the community scoring functions and validation measures are shown in Table 3. The heat maps in Figure 3 depict the rank correlations for six different networks namely, LFR (μ = 0.1), LFR (μ = 0.3), LFR (μ = 0.6), football, railway and coauthorship networks, and Table 4 presents the average values of these community scoring functions computed over all the validation measures for each of the networks. Here also, we can observe that in all the cases, the correlation is highest for permanence measure. Therefore, we can conclude that the permanence measure is more superior metric to evaluate the goodness of a detected community compared to other standard community scoring functions irrespective of any ranking schemes.

References

[1] Spearman, C., The proof and measurement of association between two things, *American Journal of Psychology* **15** (1904) 88–103.