Ricci curvature for graphs etc

TBD

(First revision, February 2, 2020)

The description is self-contained and requires minimal outside source.

1 Standard but useful notations and terminologies

Input is a connected undirected unweighted graph $G = (V(G), E(G))$ (or simply $G = (V, E)$) with the following notations:

- For a node $v \in V$, $\mathsf{Nbr}(v) = \{ u \,|\, \{v, u\} \in E \}$ denotes the set of neighbors of v, and $\mathsf{deg}(v) = |\mathsf{Nbr}(v)|$ denotes the degree of v.
- $\text{dist}_G(u, v)$ (or simply $\text{dist}(u, v)$) denote the *distance* (*i.e.*, number of edges in a shortest path) between the nodes u and v in G.

2 Ricci curvature for an edge $\{u, v\}$ (denoted by $\textsc{Ric}(u, v)$)

Consider an edge $\{u, v\}$ of G. Note that u has $\mathsf{deg}(u)$ neighbors and v has $\mathsf{deg}(v)$ neighbors, and suppose that they have ℓ neighbours in common. In other words, $\mathsf{Nbr}(u)$ and $\mathsf{Nbr}(v)$ looks likes this:

$$\mathsf{Nbr}(u) = \{ \overbrace{p_1, p_2, \dots, p_k,\ q_1, q_2, \dots, q_\ell}^{k+\ell=\mathsf{deg}(u)\geq \ell+1 \text{ nodes}} \}$$

$$\{ \underbrace{\underbrace{q_1, q_2, \dots, q_\ell}_{\substack{\ell\geq 0 \text{ common}\\ \text{neighbours}}},\ r_1, r_2, \dots, r_m}_{m+\ell=\mathsf{deg}(v)\geq \ell+1 \text{ nodes}} \} = \mathsf{Nbr}(v)$$

Note that:

- $\ell = 0$ is possible; in that case u and v have *no* common neighbors.
- **Without loss of generality, we assume that** $\mathsf{deg}(u) \leq \mathsf{deg}(v)$.

Now, assign weights (numbers) to the nodes, once from the point of view of node u and another from the point of view of node v, as follows:

- From the point of view of node u ("red" nodes, colored red for convenience only):

nodes	p_1	p_2	$\ldots$	p_k	q_1	q_2	$\ldots$	q_ℓ	u	
new names for these nodes	x_1	x_2	$\ldots$	x_k	x_{k+1}	x_{k+2}	$\ldots$	$x_{k+\ell}$	$x_{k+\ell+1} = x_{\deg(u)+1}$	
numbers	$w(x_1) = \frac{1}{\deg(u)+1}$	$w(x_2) = \frac{1}{\deg(u)+1}$	$\ldots$	$w(x_k) = \frac{1}{\deg(u)+1}$	$w(x_{k+1}) = \frac{1}{\deg(u)+1}$	$w(x_{k+2}) = \frac{1}{\deg(u)+1}$	$\ldots$	$w(x_{k+\ell}) = \frac{1}{\deg(u)+1}$	$w(x_{k+\ell}) = \frac{1}{\deg(u)+1}$	sums to 1

- From the point of view of node v ("blue" nodes, colored blue for convenience only):

nodes	q_1	q_2	$\ldots$	q_ℓ	r_1	r_2	$\ldots$	r_m	v	
new names for these nodes	y_1	y_2	$\ldots$	y_ℓ	$y_{\ell+1}$	$y_{\ell+2}$	$\ldots$	$y_{\ell+m}$	$y_{\ell+m+1} = y_{\deg(v)+1}$	
numbers	$w(y_1) = \frac{1}{\deg(v)+1}$	$w(y_2) = \frac{1}{\deg(v)+1}$	$\ldots$	$w(y_\ell) = \frac{1}{\deg(v)+1}$	$w(y_{\ell+1}) = \frac{1}{\deg(v)+1}$	$w(y_{\ell+2}) = \frac{1}{\deg(v)+1}$	$\ldots$	$w(y_{\ell+m}) = \frac{1}{\deg(v)+1}$	$w(y_{\ell+m+1}) = \frac{1}{\deg(v)+1}$	sums to 1

Then, $\textsc{Ric}(u, v)$ is *one minus* the objective value in an optimal solution to the following linear programming (LP) problem[1]:

Problem name: assign any name you wish (say Earth Mover Distance or EMD)

Reminder: $\text{dist}(x_i, y_j)$ is the distance between nodes x_i and y_j in G.

Variables: $z_{i,j}$ for $i = 1, 2, \ldots, \deg(u) + 1$ and $j = 1, 2, \ldots, \deg(v) + 1$.
Number of variables: $(\deg(u) + 1) \times (\deg(v) + 1)$.

The linear program

$$\textit{minimize} \quad \sum_{i=1}^{\deg(u)+1} \sum_{j=1}^{\deg(v)+1} \text{dist}(x_i, y_j) z_{i,j} \qquad \text{(* minimize total transportation cost *)}$$

(* the minimum objective value will be subsequently denoted by $\text{EMD}(u, v)$, thus $\textsc{Ric}(u, v) = 1 - \text{EMD}(u, v)$ *)

subject to

$$\sum_{j=1}^{\deg(v)+1} z_{i,j} = w(x_i), \quad \text{for } i = 1, 2, \ldots, \deg(u) + 1 \qquad \text{(* take from } x_i \text{ exactly as much as it has *)}$$

$$\sum_{i=1}^{\deg(u)+1} z_{i,j} = w(y_j), \quad \text{for } j = 1, 2, \ldots, \deg(v) + 1 \qquad \text{(* ship to } y_j \text{ exactly as much as it needs *)}$$

$$z_{i,j} \geq 0, \quad \text{for all } i \text{ and } j$$

3 Ricci curvature for a node u (denoted by $\textsc{Ric}(u)$)

Suppose that u has neighbors $v_1, v_2, \ldots, v_s$. Then,

$$\textsc{Ric}(u) = \frac{\textsc{Ric}(u, v_1) + \textsc{Ric}(u, v_1) + \cdots + \textsc{Ric}(u, v_s)}{s}$$

[1] Comments are enclosed by (* and *).

4 Some simple observations for $\textsc{Ric}(u, v)$

For two vectors (of numbers) $\vec{\alpha} = (\alpha_1, \ldots, \alpha_t)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_t)$, each having exactly the same number t of components, define the following quantity (usually called the total-variation (TV) distance, but names are not important):

$$\|\vec{\alpha} - \vec{\beta}\|_{TV} = \frac{|\alpha_1 - \beta_1| + \cdots + |\alpha_t - \beta_t|}{2}$$

Now, consider the following two vectors corresponding to nodes u and v:

$$\begin{array}{lcccccccccccc}
 & & p_1 & \cdots & p_k & q_1 & \cdots & q_\ell & u & r_1 & \cdots & r_m & v \\
\text{first vector (for node } u) & \vec{u} = & \Big(\frac{1}{\mathsf{deg}(u)+1} & \cdots & \frac{1}{\mathsf{deg}(u)+1} & \frac{1}{\mathsf{deg}(u)+1} & \cdots & \frac{1}{\mathsf{deg}(u)+1} & \frac{1}{\mathsf{deg}(u)+1} & 0 & \cdots & 0 & \frac{1}{\mathsf{deg}(u)+1} \Big) \\
\text{second vector (for node } v) & \vec{v} = & \Big(0 & \cdots & 0 & \frac{1}{\mathsf{deg}(v)+1} & \cdots & \frac{1}{\mathsf{deg}(v)+1} & \frac{1}{\mathsf{deg}(v)+1} & \frac{1}{\mathsf{deg}(v)+1} & \cdots & \frac{1}{\mathsf{deg}(v)+1} & \frac{1}{\mathsf{deg}(v)+1} \Big)
\end{array}$$

By straightforward calculation (remember that by our assumption $\mathsf{deg}(u) \leq \mathsf{deg}(v)$ and therefore $\frac{1}{\mathsf{deg}(u)+1} \geq \frac{1}{\mathsf{deg}(v)+1}$):

$$\begin{aligned}
\|\vec{u} - \vec{v}\|_{TV} &= \frac{1}{2} \times \left(\underbrace{\frac{1}{\mathsf{deg}(u)+1} + \cdots + \frac{1}{\mathsf{deg}(u)+1}}_{k \text{ times}} + \underbrace{\frac{1}{\mathsf{deg}(v)+1} + \cdots + \frac{1}{\mathsf{deg}(v)+1}}_{m \text{ times}} + \underbrace{\left(\frac{1}{\mathsf{deg}(u)+1} - \frac{1}{\mathsf{deg}(v)+1}\right) + \cdots + \left(\frac{1}{\mathsf{deg}(u)+1} - \frac{1}{\mathsf{deg}(v)+1}\right)}_{\ell + 2 \text{ times}} \right) \\
&= \frac{k/2}{\mathsf{deg}(u)+1} + \frac{m/2}{\mathsf{deg}(v)+1} + \frac{\ell+2}{2} \times \left(\frac{1}{\mathsf{deg}(u)+1} - \frac{1}{\mathsf{deg}(v)+1} \right) = \frac{\frac{k+\ell}{2} + 1}{\mathsf{deg}(u)+1} + \frac{\frac{m-\ell}{2} - 1}{\mathsf{deg}(v)+1}
\end{aligned}$$

Observation 1. $\textsc{Ric}(u, v) \leq 1 - \|\vec{u} - \vec{v}\|_{TV}$. *Morever, since* $1 \leq \mathrm{dist}(x_i, y_j) \leq 3$ *unless* x_i *and* y_j *are the same node, we are led to the following bound:*

$$1 - 3 \times \|\vec{u} - \vec{v}\|_{TV} \leq \textsc{Ric}(u, v) \leq 1 - \|\vec{u} - \vec{v}\|_{TV}$$

Observation 2. *Suppose that* $\mathrm{dist}(x_i, y_j)$ *is exactly equal to some fixed value* γ *for all* i *and* j. *In that case,* $\textsc{Ric}(u, v) = 1 - \gamma \times \|\vec{u} - \vec{v}\|_{TV}$.

Observation 3. *Suppose that* G *is a tree. Thus,* G *has no cycles, implying* $\ell = 0$. *Moreover,* $\mathrm{dist}(x_i, y_j) = 3$ *for all* i *and* j. *Thus, using Observation* 2 *it follows that*

$$\begin{aligned}
\textsc{Ric}(u, v) = 1 - 3 \times \|\vec{u} - \vec{v}\|_{TV} &= 1 - \frac{\frac{3k}{2} + 3}{\mathsf{deg}(u)+1} - \frac{\frac{3m}{2} - 3}{\mathsf{deg}(v)+1} \\
&= 1 - \frac{\frac{3\mathsf{deg}(u)-3}{2} + 3}{\mathsf{deg}(u)+1} - \frac{\frac{3\mathsf{deg}(v)-3}{2} - 3}{\mathsf{deg}(v)+1} = \frac{6}{\mathsf{deg}(v)+1} - 2
\end{aligned}$$

Thus, $-2 < \textsc{Ric}(u, v) \leq 0$ *and* $\textsc{Ric}(u, v)$ *can be trivially computed in linear time, i.e., in* $O(\mathsf{deg}(u) + \mathsf{deg}(v))$ *time. Note that* $\ell = 0$ *and* $\mathrm{dist}(x_i, y_j) = 3$ *for all* i *and* j *also holds for any graph* G *that does not have a cycle of* 5 *or fewer edges, so the above calculations also hold for those graphs.*

5 Some research questions for $\mathrm{Ric}(u, v)$

For notational convenience, let $\delta_1 = \mathsf{deg}(u)$ and $\delta_2 = \mathsf{deg}(v)$. The LP for computing $\mathrm{Ric}(u, v)$ involves $O(\delta_1 \times \delta_2)$ variables and $O(\delta_1 + \delta_2)$ constraints. Standard LP solvers seem to take at least quadratic time, *i.e.*, in our case $\Omega\left((\delta_1 \times \delta_2)^2\right)$ time. Observation 1 indicates that a 3-approximation of $\mathrm{Ric}(u, v)$ in linear (*i.e.*, $O(\delta_1 + \delta_2)$) time is trivial. This begs questions of the following types:

- **(Basic computational complexity questions)** Can we compute $\mathrm{Ric}(u, v)$ exactly in sub-quadratic time, *e.g.*, in $O\left((\delta_1 + \delta_2)\log(\delta_1 + \delta_2)\right)$ time or even in $O\left((\delta_1 + \delta_2)^{2-\varepsilon}\right)$ time for some constant $\varepsilon > 0$. If not, can we compute a very good (much better than 3) approximation fast?

  Note that, by Observation 3, $\mathrm{Ric}(u, v)$ can we computed in linear ($O\left(\delta_1 + \delta_2\right)$) time for any graph G that does not have a cycle of 5 or fewer edges.

- **(Relevant questions for large graphs or dynamic graphs)** Suppose that we have already computed $\mathrm{Ric}(u, v)$ and an edge (or a few edges) of the graph G gets deleted. How fast we can re-compute the new value of $\mathrm{Ric}(u, v)$? The same question applies when one or more nodes gets deleted or perhaps a collection of edges and nodes get deleted?

6 Some research questions for $\mathrm{Ric}(u)$

For notational convenience, let $\delta = \mathsf{deg}(u)$. Suppose that $\mathrm{Ric}(u, v)$ can be computed in $O(\Delta)$ time. Then, trivially, $\mathrm{Ric}(u)$ can be computed in $O(\Delta \times \delta)$ time. Can we do significantly better? The questions for large and dynamic graphs in the previous section are also applicable in this setting.

7 Some research questions for $\mathrm{Ric}(u)$ for the entire graph

Can we do better than just computing it node-by-node for the entire graph?

References

[1] No reference at all.