

# Vertex Nomination via Local Neighborhood Seeded Graph Matching

Heather G. Patsolic, Vince Lyzinski, Carey E. Priebe

Johns Hopkins University

Department of Applied Mathematics and Statistics

Baltimore, Maryland 21218-2682 USA

hgaddy1@jhu.edu, vlyzins1@jhu.edu, cep@jhu.edu



## Background/Opportunity

- The graph matching problem (GMP) aims to find a map between the vertices of one graph and the vertices of another graph which minimizes the number of edge disagreements between the two graphs.
- We assume that a portion of the bijective map is known, and utilize these known correspondences, called **seeds**, as proposed in [1].
- We are interested in a sub-problem of graph matching in which, given a vertex of interest (VOI) in one network, we seek to identify corresponding vertices in a second network.

## Challenge

- Often graphs are too large for brute-force graph matching.
- We are primarily interested in finding a particular VOI (not in matching the full networks).

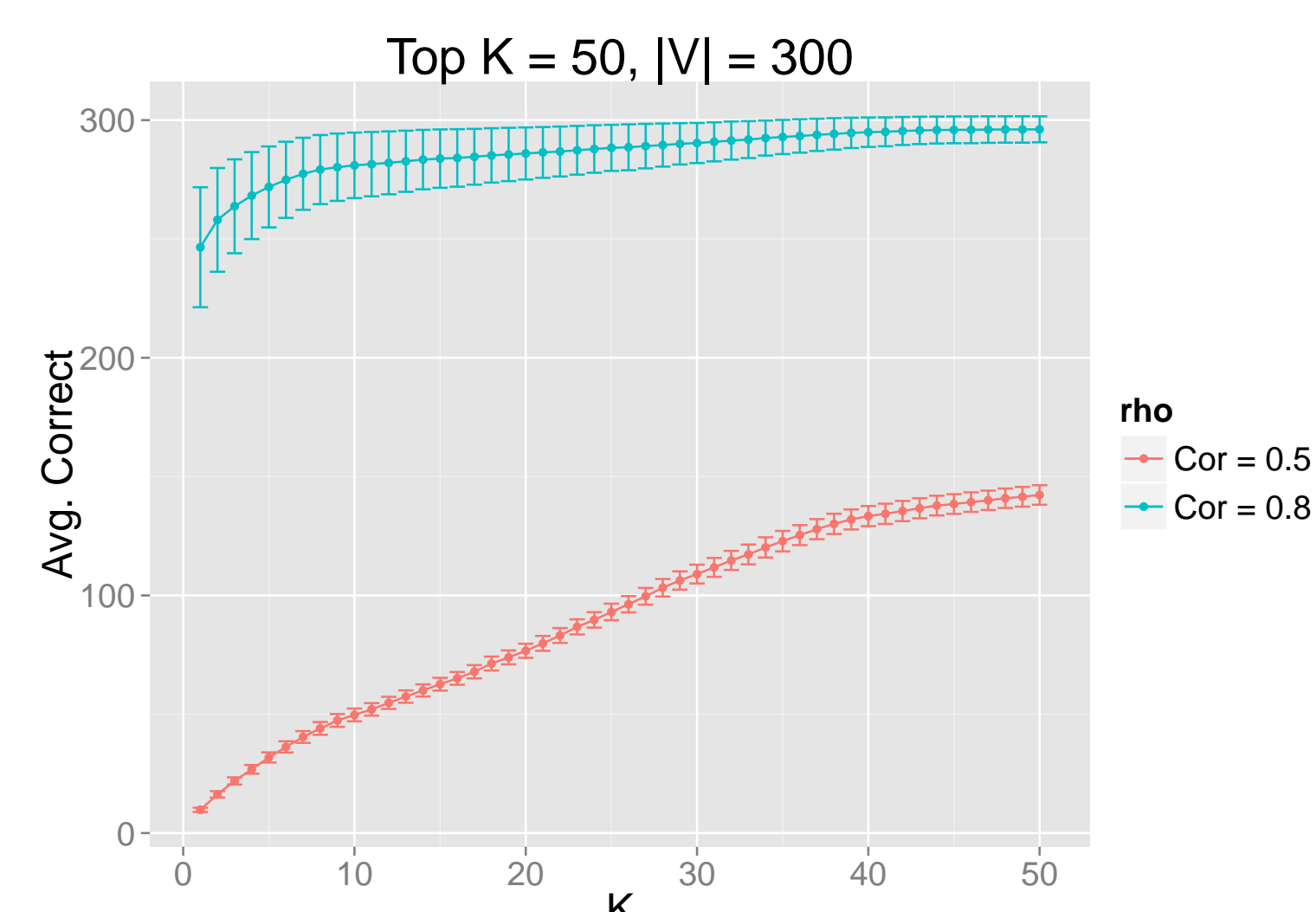
## Action

We propose the use of seeded graph matching on local neighborhoods near the VOI in order to generate a soft nomination list of vertices in the second network that are likely to correspond to the VOI in the first network. We proceed as follows.

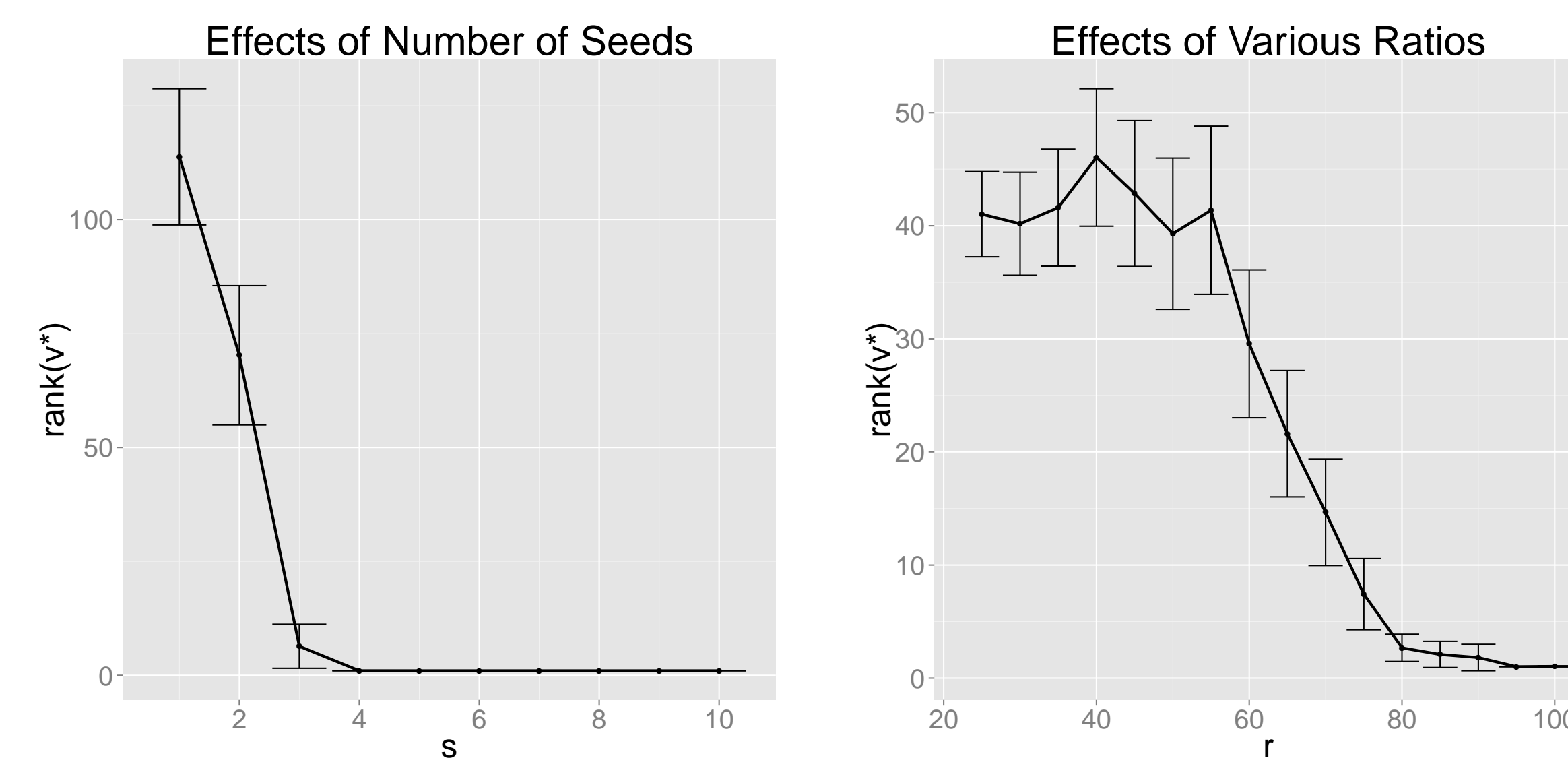
1. Identify vertices in  $h$ -neighborhood (within desired  $h$ -path) around the VOI in the first network that have verifiable corresponding vertices (seeds) in the second network.
2. Match the induced subgraphs in each network generated by the  $\ell$ -neighborhoods (for some  $\ell \geq h$ ) of these verified seeds using a modified version of the seeded graph matching algorithm presented in [1].
3. Rank the vertices of the second network in terms of the most likely matches to the original VOI,  $v^*$ . This ordered list of vertices is called the nomination list for  $v^*$ .
4. For some pre-determined  $k$ , the top- $k$  nomination list for  $v^*$  is the first  $k$  entries in the nomination list obtained in the previous step.

### 0.1 Simulations: Exploring the Effects of seeds, and differences in graph size

Demonstrate the applicability of our methodology through simulations and real data examples.



**Figure 1:** When two graphs are more highly related (correlated), fewer vertices need to be reviewed in the nomination list than when the graphs have less in common.

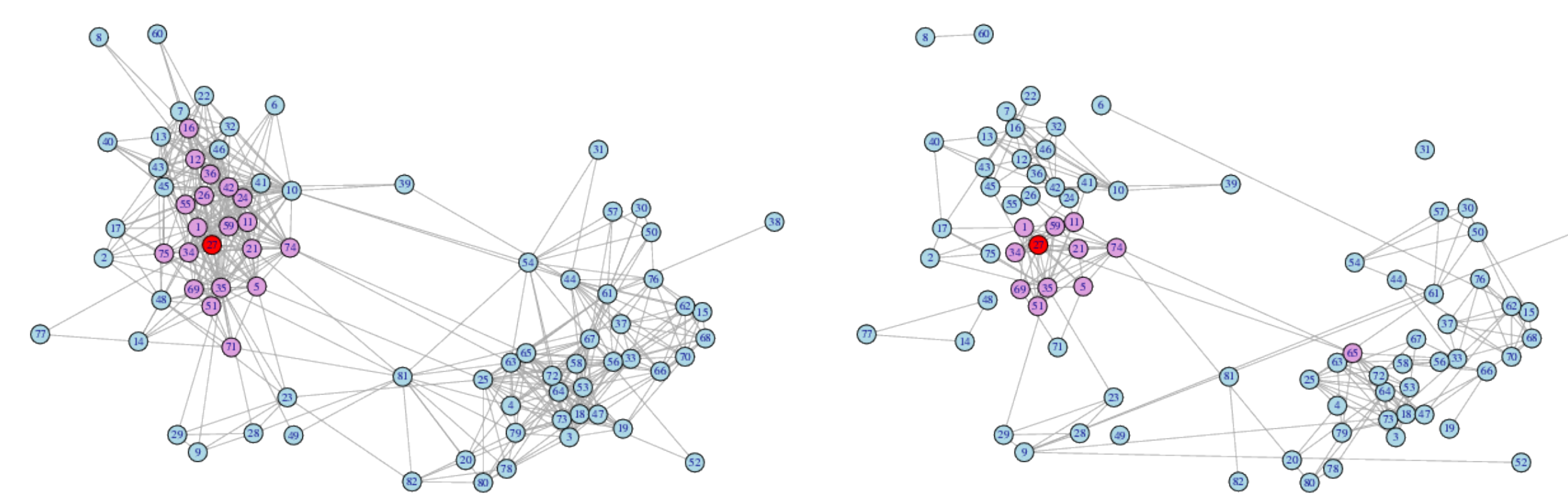


**Figure 2:** Plot of the average location of the VOI in the nomination list against: the number of seeds used in the matching (left) and the ratio of the size of the smaller graph to the larger (right).

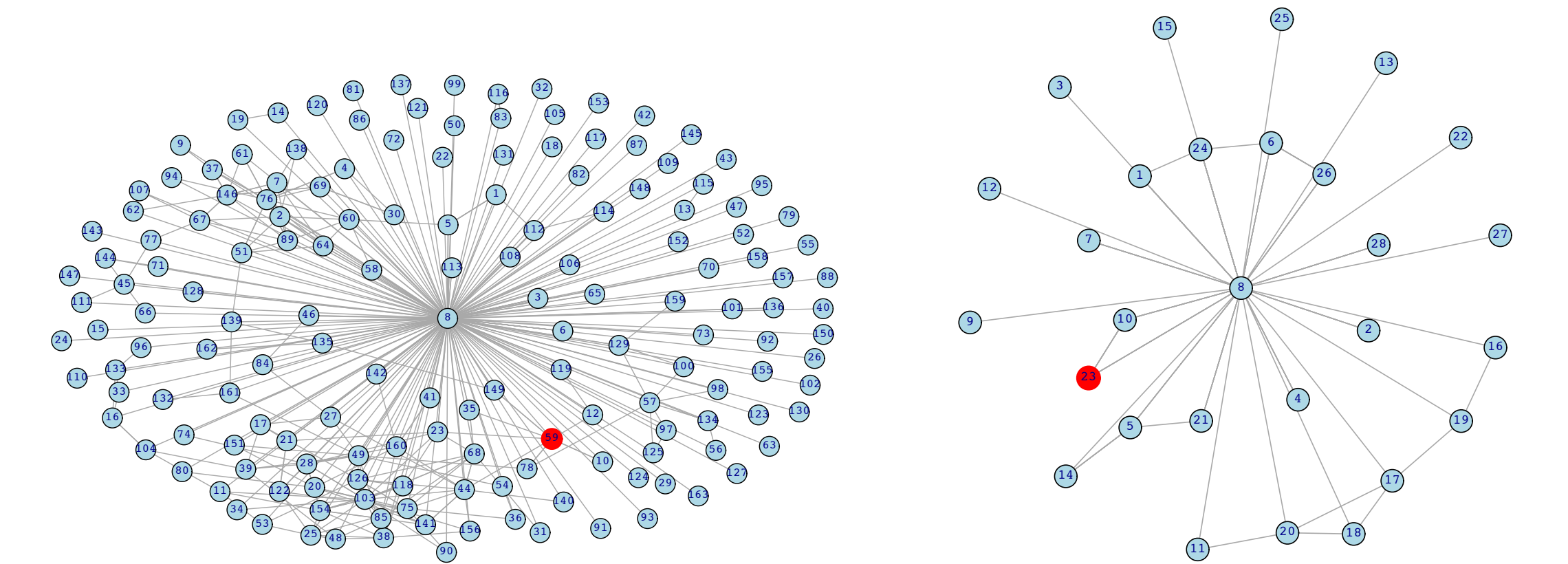
- Figure 1 shows that using our methodology (all VOI and no seeds), as the number of vertices we consider in the nomination list,  $k$ , increases, so does the number of vertices that can be correctly matched. It also shows that this matching is more accurate for graphs which are highly correlated.
- Figure 2 (left) shows that as the number of seeds increases, the location of the VOI in the nomination list decreases.
- Figure 2 (right) shows that when the graphs to match have a large discrepancy between the sizes of their vertex sets there is less accuracy in the matching.

### 0.2 Exploring real pairs of networks

We explore the effect that the number of seeds has on our methodology in two data-driven examples. The first involves a pair of high-school friendship networks as shown in Figure 3 created using data from [2], and the second is a comparison of subnetworks of Twitter and Instagram, as shown in Figure 4.

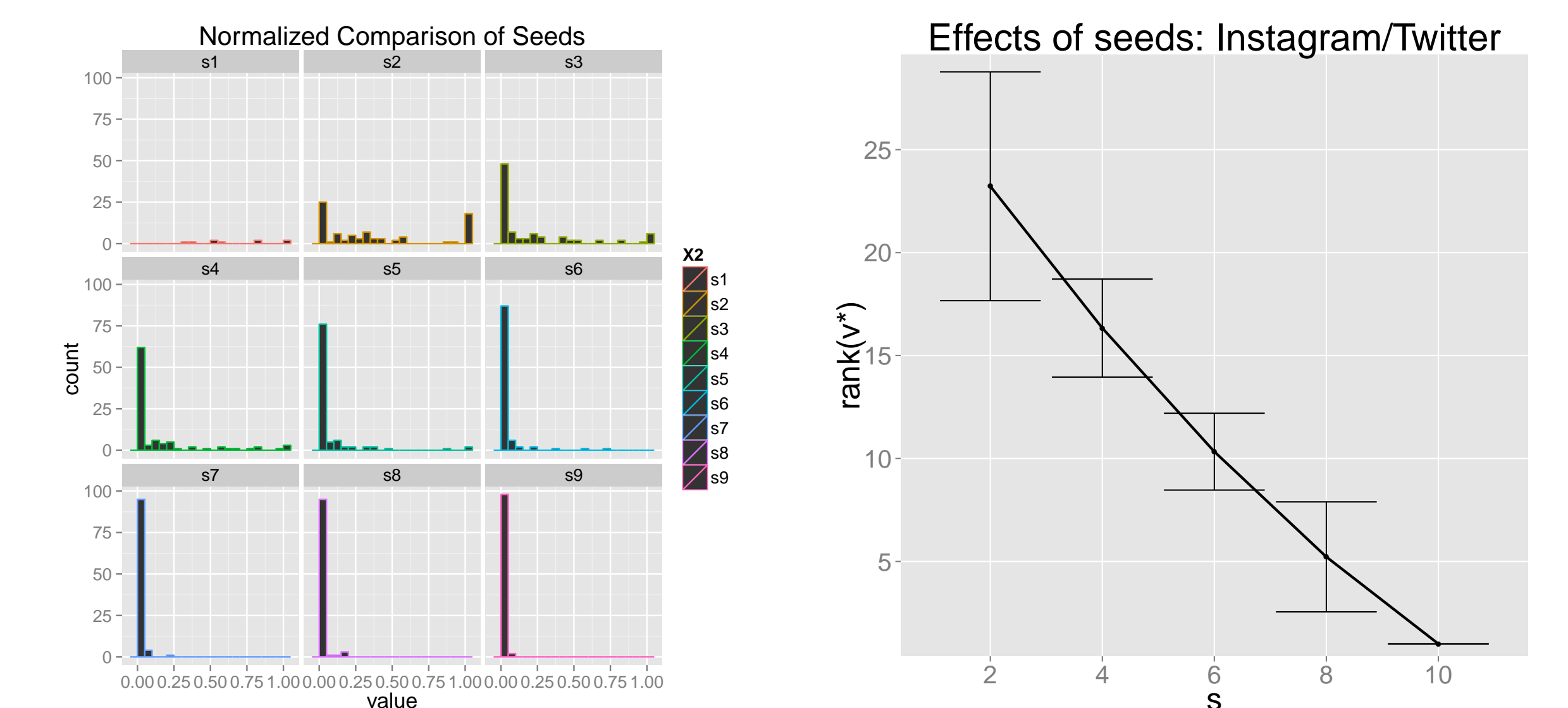


**Figure 3:** High School Facebook (left) and Friendship (right) networks.



**Figure 4:** Twitter (left) and Instagram (right) networks.

- High School Network Comparison: Setting one vertex as a fixed VOI, we sample  $s$  vertices adjacent to the VOI to use as seeds and create a histogram for each  $s \in \{1, \dots, 9\}$ , shown in Figure 5 (left), in which values of 0, 0.5, and 1 imply that the VOI was first, half-way down, and last in the nomination list, respectively. As can be seen in Figure 5 (left), by the time 3 seeds are used, our methodology is *stochastically larger* than Uniform.
- Twitter and Instagram: Letting 1 of the 11 given correspondences be the VOI, we obtain the average location of the VOI in the nomination list along with a confidence interval (as done in simulations) using an even size subset of the remaining 10 vertices. As shown in Figure 5 (right), as the number of seeds increases, the location of the VOI in the nomination list decreases.



**Figure 5:** Example of how using seeds lowers location of VOI in nomination list: pair of high school networks (left) and pair of social networks (right).

## Resolution

The provided methodology which uses seeded graph matching applied to local networks in order to generate a nomination list pertaining to a vertex of interest can be used to search larger networks when looking for a specific VOI. We demonstrate the performance of our methodology via simulations and real-data examples. This methodology is extendable to searching for multiple vertices of interest.

## References

- [1] D. E. Fishkind, S. Adali, and C. E. Priebe. Seeded graph matching. *arXiv:1209.0367*, 2012.
- [2] R. Mastrandrea, J. Fournet, and A. Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, 2015.