0.1 Comparing Aggregate Statistics of Community Structure

We begin by examining the overall statistics for the communities inferred by OSLOM using the weightings defined in Sections **??**, **??**, and **??**. The number of communities by community type is given in Table 0.1. We see that the topic- and interaction-based networks admit the most communities. The activity-based network admits the least number of communities. One advantage of the OSLOM over many other community detection algorithms is that it explicitly accounts for singleton 'communities': those nodes who do not belong to *any* extant communities. This is especially important when a network is collected via a breadth-first search, as in our network, where we begin from a seed node and then branch out. Such a search, once terminated, will result in a collection of nodes on the periphery of the network that may not belong to any community in the core.

The number of singletons by community type is also shown in Table 0.1. We see that the topic- and interaction-based communities have the most singletons, with the activity-based community dominating this measure. This result for the activity-based community is an artifact of a property of the retweet/mention weighting: many of the users in the data set (**TK: give actual number**) did not interact with each other, and thus all of their edges were zero-weighted, leading to trivial singletons.

Table 1: Number of non-singleton communities and singletons by community type: S(tructural), A(ctivity -based), T(opic-based), and I(nteraction-based).

Community Type	# of Communities	# of Singletons
S	201	308
A, Lag 1	101	951
A, Lag 2	99	600
A, Lag 3	106	611
A, Lag 4	105	668
A, Lag 5	107	632
A, Lag 6	106	642
T	289	1064
I	252	2436

Next we consider the distribution of community sizes across the community types. The complementary cumulative distribution of community sizes is given in Figure 1. Note that the axes are plotted on log-scale, and the horizontal axis begins with non-singleton communities. Thus, for a fixed community size c, Figure 1 shows the proportion of communities of size greater than c for each community type. The largest communities for the structural, activity-based, topic-based, and interaction-based networks are 198, 358, 338, and 811 respectively.

Next, we compare the number of users which belong to more than one community. Figure 2 shows the number of users belonging to 2, 3, or 4 communities. We see that as the number of mixed membership communities increases, the number of users with that number of mixed memberships decreases. This is especially true for the activity-based community **TK: speculate on what this means? Or save for the results section?**. In addition, **TK: mention the 5, 6, and 7 cases, not included in the**

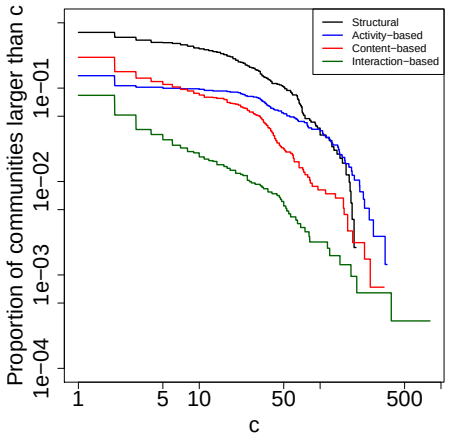

Figure 1: The proportion of communities greater than c in size, across the different community types. Note the logarithmic scale on the horizontal and vertical axes.

figure. This corresponded to **TK: investigate which users are the high-overlap and what communities they belong to.**

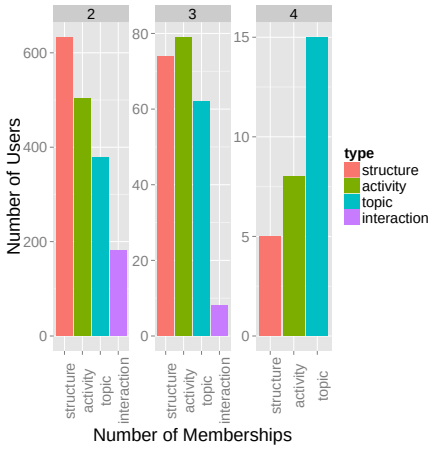

Figure 2: The number of users belonging to 2, 3, or 4 communities, by community type.