Supplementary material

Title: Impact of degree truncation on the spread of a contagious process on networks

Supplementary Content 1:

This content provides a more in-depth description of: (i) the impact of contact truncation on structural network properties; and (ii) the impact of network properties on spreading outcomes.

The impact of contact truncation on structural network properties

Degree distribution and assortativity. While the impact of FCD on the *network degree distribution* p_k is almost always to reduce its mean and variance, its precise effect depends on both the first and second moment of the degree distribution and on the ratio of k_{fc} to the mean degree μ_k. Loss of edges in high-variance networks may, however, be offset by degree-assortativity (Kossinets, 2006), often quantified by the Pearson correlation coefficient of degrees of connected nodes: $r = \frac{\Sigma_{xy}(e_{xy} - a_x b_y)}{\sigma_a \sigma_b}$, where e_{xy} is the fraction of all edges that join nodes of degree x and y, a_x and b_y are the fraction of edges that start and end, respectively, at nodes of degree x or y, respectively, and $\sigma_a$and $\sigma_b$are the standard deviations of distributions of a_x and b_y (Newman, 2003b). If the network is degree-disassortative, such as the scale-free Barabási-Albert network where $p_k \sim k^{-\gamma}$ and $2 < \gamma < 3$ (Barabási & Albert, 1999), then edges that might be censored by the adjacent high-degree node are less likely to also be censored by the adjacent low-degree node, and thus dropped entirely in the truncated network (Vázquez & Moreno, 2003). Degree-assortative, high-variance networks are thus likely to see the greatest change in p_k; human contact networks are typically somewhat degree-assortative, and while communication contacts have fat-tailed degree distributions with high variance, physical contact networks are more degree-homogeneous (Onnela et al., 2007a; Onnela et al., 2007b; Salathé et al., 2010). The level of degree-assortativity in a network is not itself systematically affected by FCD, so long as edges are dropped without regard to the strength of each connection (Kossinets, 2006; Lee et al., 2006). However, if individuals are more likely to report stronger connections, and ties between individuals of similar degree are more likely to be strong – which is suggested by the combination of findings that homophilous ties are more likely to be transitive (Louch, 2000; Marsden, 1987) and those with greater transitivity (Onnela et al., 2007b) tend to be stronger – then FCD might be expected to artificially inflate r.

Clustering. Local clustering can be measured in at least two different ways: (i) *Triadic clustering*: the mean of local clustering coefficient C_i, where C_i is the ratio of the number of ties present between all neighbors of node i and $k_i(k_i - 1)/2$, the number of pairs of neighbors of i (Watts & Strogatz, 1998); (ii) *Focal clustering*: the level of global triadic closure, that is the ratio of triangles – where (u, v), (u, w) and (v, w) are all present – to paths of length two, i.e., if (u, v) and (v, w) exist, they form a path of length two (Newman, 2010). Clustering may also occur at higher levels of aggregation in the network, for example in the presence of network communities where, loosely speaking, the density of edges within a set of nodes belonging to a community is higher than the average density of edges across the whole graph (Fortunato, 2010; Porter et al., 2009). One way to quantify this community-level clustering is by modularity, $Q = \sum_r (e_{rr} - a_r^2)$, where e_{rr} is the proportion of edges in the network that connect nodes in community r to other nodes in community r and a_r is the proportion of ends of edges that are attached to nodes in community r (Newman, 2006). The value of modularity can be normalized using the degree distribution of the network as $Q_n = Q/\left[1 - \sum_r \left(k_i k_j/2m\right)\delta(c_i, c_j)/2m\right]$, m is the number of edges in the network and $\delta(c_i, c_j)$ is equal to one if $c_i = c_j$ and zero otherwise. This normalization makes modularity values more readily comparable across networks (Newman, 2010).

When truncation is unweighted, we expect FCD to reduce clustering at the triadic and community levels as it effectively results in random edge removal. When truncation is weighted, however, FCD might lead to an increase in clustering: if within-cluster edges are stronger than others, they are more likely to be preserved.

Path lengths. In removing ties, unweighted FCD will reduce the fractional size of the *largest connected component* (LCC), S_{LCC}, and will often increase the *average path length between nodes of the LCC*, ℓ_{LCC}, insofar as the increased length between some pairs of nodes due to loss of edges is not offset by reductions in length due to peripheral nodes being dropped altogether from the LCC. These results are seen asymptotically for random and power law graphs (Fernholz & Ramachandran, 2007), and via simulation of edge removal on empirical networks (Onnela et al., 2007a). If FCD is weighted, this second factor will be stronger, as peripherally (weakly) connected nodes are preferentially dropped from the LCC. In a network with a dense core, the S_{LCC} is likely to be better preserved in a degree-disassortative than in a degree-assortative network under FCD – due to the lower probability of ties within the core being dropped from both ends (Kossinets, 2006). This effect will be more pronounced if the ties within this core are also stronger than other ties, and thus more likely to be preserved.

While the above discussion considers structurally shortest paths between pairs of nodes, random spreading processes rarely follow shortest paths between any two nodes i and j. Because of this, the length of the shortest path between i and j in a fully observed network typically underestimates the length of the path taken by a spreading process. Partial observation of the network, such as that induced by degree truncation, inflates the lengths of the observed shortest paths, but does of course not alter the length of the actual unobserved paths taken by the spreading process. For this reason, perhaps somewhat paradoxically, shortest paths inferred from partially observed networks can provide more accurate predictions of the path lengths taken by spreading processes than those based on fully observed networks (Onnela & Christakis, 2012).

The impact of structural network properties on spreading processes

Degree distribution and assortativity. In a network setting, R_0 can be viewed as the average number of edges through which an individual infects their neighbors across the whole period of their infectiousness, if all their neighbors are susceptible. The probability of infection for each node, λ, can be conceptualized in terms of their degree and their neighbors' infection statuses. In a degree-homogenous network, a degree infectivity epidemic will probabilistically take off if the infection probabilities across the degree distribution $\lambda(k) \equiv \langle k \rangle \geq 1$, where $\langle k \rangle$ is the first moment (mean) of the degree distribution of all nodes in the network. In degree-heterogeneous networks, the likelihood of epidemic take-off becomes a function of the first and second moments of the degree distribution (Pastor-Satorras & Vespignani, 2002), such that higher degree heterogeneity increases R_0.Similarly, higher degree-assortativity increases the chances of epidemic take-off. The probabilistic threshold for epidemic take-off has a lower-bound of $\langle k_{nn} \rangle$, the average degree of nearest neighbors, which is also the driver of both degenerate results: $\langle k_{nn} \rangle = \langle k \rangle$ in a homogeneous network and $\langle k_{nn} \rangle \to \infty$ in an infinitely large scale-free network (Boguñá et al., 2003). This is intuitive, since the number of one's neighbors bounds the number of infections one can generate. Degree-assortativity leads to faster take-off, but a lower attack rate, conditional on the number of nodes and ties within a network (Gupta et al., 1989). This result arises from a dense core of high-degree nodes in which infection is rapidly passed, in combination with longer paths to peripheral, low-degree nodes where chains of infection are more likely to die out. On scale-free networks, an epidemic will grow at a power law rate, such that early in the epidemic infection levels will be greater than is predicted by homogeneous models, in which growth rates are exponential (Vazquez, 2006).

Clustering. The most straightforward effect of triadic clustering, for a given degree distribution, is to reduce the average number of infections each infected person causes. This reduction is due to newly-infected individuals having fewer susceptible neighbors: the contact who infected you is likely also have had the opportunity to infect your other contacts (Keeling, 2005; Miller, 2009; Molina & Stone, 2012). This does not strictly imply a lower R_0, since R_0 refers to a completely susceptible population, however this phenomenon increases the epidemic threshold in the same manner that a fall in R_0 would (Molina & Stone, 2012). Similarly, the epidemic growth rate r_0 is somewhat slowed by this reduction in the proportion of susceptible alters (Eames, 2008).

In many networks, e.g. Erdős–Rényi graphs (Erdős & Rényi, 1959), for a given network density, increased clustering also leads to a smaller S_{LCC}, which necessarily reduces the maximum possible attack rate (Newman, 2003a). However, within the LCC clustering increases the density of the network (Serrano & Boguná, 2006), providing more local pathways from an infected to a susceptible individual. This reduces the protective effect of any alters who have recovered without infecting an ego, and thus some simulations have found clustering increases the attack rate A (Keeling, 2005; Newman, 2003a).

Overall, cliques alone appear to have marginal effects on epidemic dynamics, however the processes which drive clique formation – such as homophily by nodal attributes or geographic proximity – lead to networks displaying clustering that also contain other topological features – such as degree-assortativity or heterogeneity – which do significantly affect epidemics, leading to processes on clustered networks looking very different from those on non-clustered ones

(Badham & Stocker, 2010; Molina & Stone, 2012; Volz et al., 2011). Broader community structure in networks acts in much the same fashion as cliques, reducing r_0 due to limited capacity to pass infection from one community to the next (Salathé & Jones, 2010); although epidemics are unhindered, or even sped up, by inter-community ties when overlapping, rather than distinctly separated, communities are built into networks (Reid & Hurley, 2011).

References

Badham, J., & Stocker, R. (2010). The impact of network clustering and assortativity on epidemic behaviour. Theoretical Population Biology, 77(1), 71-75.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.

Boguñá, M., Pastor-Satorras, R., & Vespignani, A. (2003). Absence of epidemic threshold in scale-free networks with degree correlations. [Research Support, Non-U.S. Gov't]. Physical Review Letters, 90(2), 028701.

Eames, K. T. (2008). Modelling disease spread through random and regular contacts in clustered populations. Theoretical Population Biology, 73(1), 104-111.

Erdős, P., & Rényi, A. (1959). On random graphs I. Publicationes Mathematicae, 6, 290-297.

Fernholz, D., & Ramachandran, V. (2007). The diameter of sparse random graphs. Random Structures & Algorithms, 31(4), 482-516.

Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3), 75-174.

Gupta, S., Anderson, R. M., & May, R. M. (1989). Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS, 3(12), 807-818.

Keeling, M. (2005). The implications of network structure for epidemic dynamics. Theoretical Population Biology, 67(1), 1-8.

Kossinets, G. (2006). Effects of missing data in social networks. Social networks, 28(3), 247-268.

Lee, S. H., Kim, P.-J., & Jeong, H. (2006). Statistical properties of sampled networks. Physical Review E, 73(1), 016102.

Louch, H. (2000). Personal network integration: transitivity and homophily in strong-tie relations. Social Networks, 22(1), 45-64.

Marsden, P. V. (1987). Core discussion networks of Americans. American Sociological Review, 122-131.

Miller, J. C. (2009). Spread of infectious disease through clustered populations. Journal of The Royal Society Interface, 6(41), 1121-1134.

Molina, C., & Stone, L. (2012). Modelling the spread of diseases in clustered networks. Journal of Theoretical Biology, 315, 110-118.

Newman, M. E. (2003a). Properties of highly clustered networks. Physical Review E, 68(2), 026121.

Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577-8582.

Newman, M. E. (2010). Networks: An Introduction: Oxford University Press.

Newman, M. E. J. (2003b). Mixing patterns in networks. Physical Review E, 67(2), 1-13. doi: 10.1103/PhysRevE.67.026126

Onnela, J.-P., & Christakis, N. A. (2012). Spreading paths in partially observed social networks. Physical Review E, 85(3), 036106.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., . . . Kertész, J. (2007a). Analysis of a large-scale weighted network of one-to-one human communication. New Journal of Physics, 9(6), 179.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., . . . Barabási, A.-L. (2007b). Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences, 104(18), 7332-7336.

Pastor-Satorras, R., & Vespignani, A. (2002). Immunization of complex networks. Physical Review E, 65(3), 036104.

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. Notices of the AMS, 56(9), 1082-1097.

Reid, F., & Hurley, N. (2011). Diffusion in networks with overlapping community structure. Paper presented at the Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on.

Salathé, M., & Jones, J. H. (2010). Dynamics and control of diseases in networks with community structure. PLoS Computational Biology, 6(4), e1000736.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences, 107(51), 22020-22025.

Serrano, M. Á., & Boguná, M. (2006). Percolation and epidemic thresholds in clustered networks. Physical Review Letters, 97(8), 088701.

Vazquez, A. (2006). Spreading dynamics on small-world networks with connectivity fluctuations and correlations. Physical Review E, 74(5), 056101.

Vázquez, A., & Moreno, Y. (2003). Resilience to damage of graphs with degree correlations. Physical Review E, 67(1), 015101.

Volz, E. M., Miller, J. C., Galvani, A., & Meyers, L. A. (2011). Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. PLoS Computational Biology, 7(6), e1002042.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’networks. Nature, 393(6684), 440-442.

Supplementary Table 1: Descriptive statistics for the calibrated network graphs (mean and interquartile range)

A. <u>Mean degree</u>

		Not truncated		Truncated at $2\langle k\rangle$		Truncated at $\langle k\rangle$		Truncated at $0.5\langle k\rangle$	
Karnataka villages		8.39	[7.84 - 8.97]	7.21	[6.72 - 7.60]	5.54	[4.77 - 5.65]	3.90	[2.78 - 3.95]
Synthetic networks:									
Degree-Assortative	r = 0.283	7.86	[7.86 - 7.86]	7.68	[7.67 - 7.68]	5.74	[5.72 - 5.76]	3.22	[3.20 - 3.24]
	r = 0.421	7.86	[7.86 - 7.86]	7.64	[7.63 - 7.65]	5.67	[5.65 - 5.69]	3.16	[3.13 - 3.18]
	r = 0.797	7.86	[7.86 - 7.86]	7.54	[7.53 - 7.55]	5.40	[5.38 - 5.42]	2.93	[2.91 - 2.95]
Triadic Clustering	c = 0.249	7.75	[7.75 - 7.75]	7.40	[7.39 - 7.42]	5.56	[5.53 - 5.58]	3.12	[3.10 - 3.13]
	c = 0.284	7.75	[7.75 - 7.75]	7.39	[7.36 - 7.40]	5.55	[5.52 - 5.57]	3.19	[3.17 - 3.20]
	c = 0.353	7.75	[7.75 - 7.75]	7.31	[7.29 - 7.33]	5.51	[5.48 - 5.53]	3.32	[3.30 - 3.33]
Focal Clustering	t = 0.163	7.95	[7.95 - 7.95]	6.84	[6.78 - 6.88]	4.49	[4.46 - 4.54]	2.57	[2.54 - 2.59]
	t = 0.249	7.95	[7.95 - 7.95]	6.29	[6.17 - 6.37]	4.07	[4.00 - 4.12]	2.32	[2.28 - 2.35]
	t = 0.326	7.95	[7.95 - 7.95]	5.84	[5.73 - 5.92]	3.76	[3.67 - 3.83]	2.15	[2.11 - 2.20]
Power-Law	$\gamma = 3$	7.78	[7.66 - 7.83]	6.58	[6.50 - 6.63]	4.70	[4.66 - 4.74]	2.89	[2.87 - 2.91]
	$\gamma = 2.5$	7.40	[7.04 - 7.55]	6.22	[5.97 - 6.33]	4.60	[4.56 - 4.65]	2.91	[2.89 - 2.93]
	$\gamma = 2$	6.18	[5.89 - 6.46]	4.78	[4.44 - 5.02]	4.00	[3.51 - 4.18]	2.88	[2.85 - 2.91]

B. Degree-assortativity

		Not truncated		Truncated at $2\langle k \rangle$		Truncated at $\langle k \rangle$		Truncated at $0.5\langle k \rangle$	
Karnataka villages		0.33	[0.30 - 0.37]	0.23	[0.20 - 0.25]	0.11	[0.09 - 0.13]	0.02	[-0.02 - 0.05]
Synthetic networks defined by:									
Degree-Assortative	r = 0.283	0.28	[0.28 - 0.28]	0.25	[0.25 - 0.26]	-0.02	[-0.03 - -0.01]	-0.19	[-0.20 - -0.18]
	r = 0.421	0.42	[0.42 - 0.42]	0.38	[0.37 - 0.38]	-0.00	[-0.01 - 0.01]	-0.19	[-0.20 - -0.17]
	r = 0.797	0.80	[0.80 - 0.80]	0.69	[0.68 - 0.69]	-0.00	[-0.02 - 0.01]	-0.20	[-0.21 - -0.18]
Triadic Clustering	c = 0.249	-0.05	[-0.06 - -0.04]	-0.10	[-0.11 - -0.09]	-0.16	[-0.17 - -0.15]	-0.25	[-0.27 - -0.24]
	c = 0.284	-0.05	[-0.06 - -0.04]	-0.10	[-0.11 - -0.09]	-0.17	[-0.18 - -0.16]	-0.26	[-0.27 - -0.25]
	c = 0.353	-0.06	[-0.07 - -0.05]	-0.11	[-0.12 - -0.10]	-0.18	[-0.19 - -0.17]	-0.27	[-0.28 - -0.26]
Focal Clustering	t = 0.163	0.26	[0.23 - 0.29]	0.11	[0.09 - 0.12]	-0.07	[-0.08 - -0.06]	-0.18	[-0.20 - -0.17]
	t = 0.249	0.50	[0.46 - 0.55]	0.12	[0.11 - 0.14]	-0.10	[-0.11 - -0.08]	-0.20	[-0.22 - -0.19]
	t = 0.326	0.68	[0.65 - 0.72]	0.08	[0.07 - 0.10]	-0.14	[-0.15 - -0.13]	-0.23	[-0.25 - -0.21]
Power-Law	$\gamma = 3$	-0.04	[-0.06 - -0.03]	-0.11	[-0.13 - -0.09]	-0.12	[-0.15 - -0.10]	-0.14	[-0.18 - -0.10]
	$\gamma = 2.5$	-0.10	[-0.13 - -0.08]	-0.14	[-0.16 - -0.12]	-0.14	[-0.16 - -0.11]	-0.14	[-0.16 - -0.12]
	$\gamma = 2$	-0.22	[-0.24 - -0.20]	-0.24	[-0.26 - -0.21]	-0.23	[-0.26 - -0.21]	-0.22	[-0.25 - -0.20]

C. Modularity

		Not truncated		Truncated at $2\langle k \rangle$		Truncated at $\langle k \rangle$		Truncated at $0.5\langle k \rangle$	
Karnataka villages		0.79	[0.77 - 0.82]	0.81	[0.79 - 0.84]	0.84	[0.82 - 0.86]	0.87	[0.84 - 0.90]
Synthetic networks defined by:									
Degree-Assortative	r = 0.283	0.29	[0.29 - 0.29]	0.30	[0.30 - 0.30]	0.40	[0.40 - 0.41]	0.66	[0.65 - 0.66]
	r = 0.421	0.28	[0.28 - 0.29]	0.30	[0.30 - 0.30]	0.41	[0.40 - 0.41]	0.66	[0.66 - 0.67]
	r = 0.797	0.28	[0.28 - 0.28]	0.30	[0.30 - 0.30]	0.44	[0.43 - 0.45]	0.71	[0.71 - 0.72]
Triadic Clustering	c = 0.249	0.46	[0.45 - 0.46]	0.46	[0.45 - 0.46]	0.48	[0.48 - 0.49]	0.68	[0.67 - 0.68]
	c = 0.284	0.47	[0.47 - 0.48]	0.47	[0.47 - 0.48]	0.49	[0.49 - 0.50]	0.67	[0.67 - 0.68]
	c = 0.353	0.50	[0.49 - 0.50]	0.50	[0.49 - 0.50]	0.52	[0.51 - 0.52]	0.66	[0.66 - 0.67]
Focal Clustering	t = 0.163	0.66	[0.65 - 0.67]	0.62	[0.61 - 0.63]	0.60	[0.59 - 0.60]	0.76	[0.76 - 0.77]
	t = 0.249	0.82	[0.81 - 0.83]	0.78	[0.77 - 0.79]	0.72	[0.72 - 0.74]	0.81	[0.81 - 0.82]
	t = 0.326	0.90	[0.89 - 0.91]	0.87	[0.86 - 0.89]	0.83	[0.81 - 0.84]	0.86	[0.85 - 0.87]
Power-Law	$\gamma = 3$	0.36	[0.36 - 0.36]	0.32	[0.31 - 0.32]	0.43	[0.43 - 0.44]	0.68	[0.67 - 0.69]
	$\gamma = 2.5$	0.36	[0.35 - 0.36]	0.34	[0.33 - 0.35]	0.45	[0.45 - 0.46]	0.68	[0.67 - 0.68]
	$\gamma = 2$	0.37	[0.36 - 0.38]	0.43	[0.41 - 0.45]	0.50	[0.49 - 0.56]	0.68	[0.67 - 0.68]

D. Triadic clustering coefficient

		Not truncated		Truncated at $2\langle k\rangle$		Truncated at $\langle k\rangle$		Truncated at $0.5\langle k\rangle$	
Karnataka villages		0.64	[0.63 - 0.66]	0.60	[0.57 - 0.61]	0.50	[0.48 - 0.51]	0.34	[0.27 - 0.37]
Synthetic networks defined by:									
Degree-Assortative	r = 0.283	0.01	[0.01 - 0.01]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.01]	0.00	[0.00 - 0.00]
	r = 0.421	0.01	[0.01 - 0.01]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.01]	0.00	[0.00 - 0.00]
	r = 0.797	0.01	[0.01 - 0.01]	0.01	[0.01 - 0.01]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.00]
Triadic Clustering	c = 0.249	0.29	[0.29 - 0.30]	0.26	[0.26 - 0.26]	0.13	[0.12 - 0.13]	0.03	[0.03 - 0.04]
	c = 0.284	0.34	[0.34 - 0.34]	0.30	[0.29 - 0.30]	0.15	[0.15 - 0.16]	0.04	[0.04 - 0.05]
	c = 0.353	0.43	[0.43 - 0.43]	0.37	[0.36 - 0.37]	0.20	[0.19 - 0.20]	0.07	[0.06 - 0.07]
Focal Clustering	t = 0.163	0.37	[0.37 - 0.38]	0.28	[0.27 - 0.28]	0.12	[0.12 - 0.13]	0.04	[0.04 - 0.05]
	t = 0.249	0.43	[0.42 - 0.44]	0.30	[0.29 - 0.31]	0.15	[0.13 - 0.15]	0.06	[0.05 - 0.06]
	t = 0.326	0.45	[0.44 - 0.46]	0.31	[0.30 - 0.32]	0.16	[0.15 - 0.17]	0.06	[0.05 - 0.07]
Power-Law	$\gamma = 3$	0.04	[0.03 - 0.05]	0.02	[0.02 - 0.02]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.01]
	$\gamma = 2.5$	0.09	[0.07 - 0.13]	0.04	[0.03 - 0.05]	0.02	[0.02 - 0.03]	0.01	[0.01 - 0.01]
	$\gamma = 2$	0.21	[0.19 - 0.22]	0.05	[0.04 - 0.06]	0.03	[0.03 - 0.03]	0.02	[0.01 - 0.02]

E. Focal clustering coefficient

		Not truncated		Truncated at $2\langle k\rangle$		Truncated at $\langle k\rangle$		Truncated at $0.5\langle k\rangle$	
Karnataka villages		0.19	[0.17 - 0.21]	0.18	[0.16 - 0.19]	0.16	[0.15 - 0.17]	0.11	[0.08 - 0.12]
Synthetic networks defined by:									
Degree-Assortative	r = 0.283	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]
	r = 0.421	0.01	[0.00 - 0.01]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]
	r = 0.797	0.01	[0.01 - 0.01]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]
Triadic Clustering	c = 0.249	0.07	[0.07 - 0.07]	0.06	[0.06 - 0.06]	0.03	[0.03 - 0.03]	0.01	[0.01 - 0.01]
	c = 0.284	0.08	[0.08 - 0.08]	0.07	[0.07 - 0.07]	0.03	[0.03 - 0.03]	0.01	[0.01 - 0.01]
	c = 0.353	0.09	[0.08 - 0.09]	0.07	[0.07 - 0.07]	0.04	[0.04 - 0.04]	0.01	[0.01 - 0.01]
Focal Clustering	t = 0.163	0.16	[0.16 - 0.16]	0.11	[0.10 - 0.11]	0.05	[0.04 - 0.05]	0.01	[0.01 - 0.02]
	t = 0.249	0.25	[0.25 - 0.25]	0.14	[0.13 - 0.14]	0.06	[0.06 - 0.06]	0.02	[0.02 - 0.02]
	t = 0.326	0.33	[0.33 - 0.33]	0.15	[0.15 - 0.16]	0.07	[0.06 - 0.07]	0.02	[0.02 - 0.03]
Power-Law	γ = 3	0.02	[0.02 - 0.02]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]
	γ = 2.5	0.03	[0.02 - 0.03]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]
	γ = 2	0.04	[0.04 - 0.05]	0.01	[0.01 - 0.01]	0.00	[0.00 - 0.00]	0.00	[0.00 - 0.00]

F. Average shortest path in Largest Connected Component

		Not truncated		Truncated at $2\langle k\rangle$		Truncated at $\langle k\rangle$		Truncated at $0.5\langle k\rangle$	
Karnataka villages		4.10	[3.89 - 4.36]	4.43	[4.19 - 4.68]	5.30	[5.00 - 5.82]	7.09	[6.56 - 9.23]
Synthetic networks defined by:									
Degree-Assortative	r = 0.283	3.61	[3.61 - 3.62]	3.65	[3.65 - 3.65]	4.17	[4.16 - 4.18]	6.17	[6.13 - 6.23]
	r = 0.421	3.65	[3.65 - 3.65]	3.69	[3.69 - 3.69]	4.22	[4.21 - 4.23]	6.36	[6.29 - 6.41]
	r = 0.797	3.88	[3.87 - 3.88]	3.91	[3.90 - 3.91]	4.47	[4.47 - 4.48]	7.36	[7.28 - 7.46]
Triadic Clustering	c = 0.249	3.71	[3.70 - 3.72]	3.78	[3.77 - 3.79]	4.22	[4.20 - 4.23]	6.35	[6.28 - 6.42]
	c = 0.284	3.70	[3.70 - 3.72]	3.78	[3.77 - 3.79]	4.21	[4.20 - 4.23]	6.11	[6.05 - 6.17]
	c = 0.353	3.69	[3.68 - 3.70]	3.78	[3.77 - 3.79]	4.20	[4.18 - 4.22]	5.75	[5.70 - 5.80]
Focal Clustering	t = 0.163	4.09	[4.07 - 4.12]	4.21	[4.18 - 4.23]	4.91	[4.88 - 4.94]	7.94	[7.84 - 8.07]
	t = 0.249	4.61	[4.56 - 4.66]	4.73	[4.68 - 4.78]	5.39	[5.34 - 5.45]	8.33	[8.26 - 8.47]
	t = 0.326	5.23	[5.10 - 5.39]	5.34	[5.20 - 5.51]	5.98	[5.83 - 6.17]	8.85	[8.60 - 9.14]
Power-Law	γ = 3	3.35	[3.30 - 3.38]	3.61	[3.56 - 3.64]	4.25	[4.18 - 4.30]	6.34	[6.12 - 6.51]
	γ = 2.5	3.16	[3.09 - 3.23]	3.43	[3.36 - 3.51]	3.93	[3.79 - 4.06]	5.52	[5.22 - 5.80]
	γ = 2	3.07	[3.03 - 3.10]	3.50	[3.45 - 3.54]	3.85	[3.79 - 3.93]	4.70	[4.59 - 4.83]

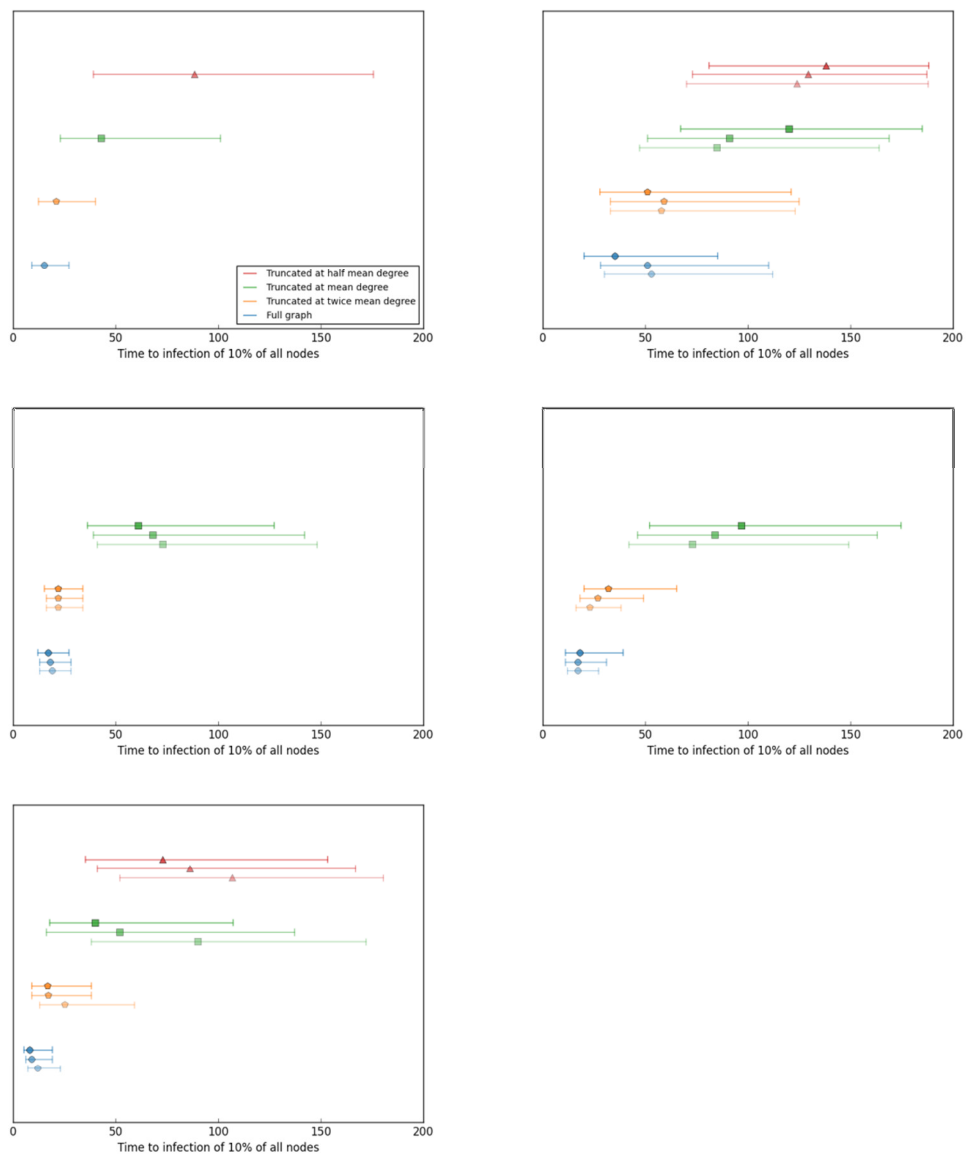
$\langle k\rangle$: Mean degree of nodes in a given graph. For definitions of r, c, t, γ and δ and how they define each synthetic network type, please see main text of paper.

Supplementary Table 2: Percentage of epidemic simulation runs infecting at least 10% of the population

		Not truncated	Truncated at $2\langle k\rangle$	Truncated at $\langle k\rangle$	Truncated at $0.5\langle k\rangle$
Karnataka villages		99.5	99.3	90.4	11.9
Synthetic networks defined by:					
Degree-Assortative	r = 0.283	91.1	90.1	76.0	11.7
	r = 0.421	89.9	88.9	69.3	8.7
	r = 0.797	89.1	82.6	26.7	1.0
Triadic Clustering	c = 0.249	99.8	99.8	87.3	0.0
	c = 0.284	99.9	99.8	92.1	0.0
	c = 0.353	99.8	99.8	95.9	0.0
Focal Clustering	t = 0.163	99.6	99.4	55.6	0.0
	t = 0.249	98.9	98.3	66.0	0.0
	t = 0.326	97.5	96.4	66.7	0.0
Power-Law	$\gamma = 3$	98.6	92.1	43.6	9.7
	$\gamma = 2.5$	98.9	95.1	51.9	15.0
	$\gamma = 2$	97.5	89.1	56.3	23.8

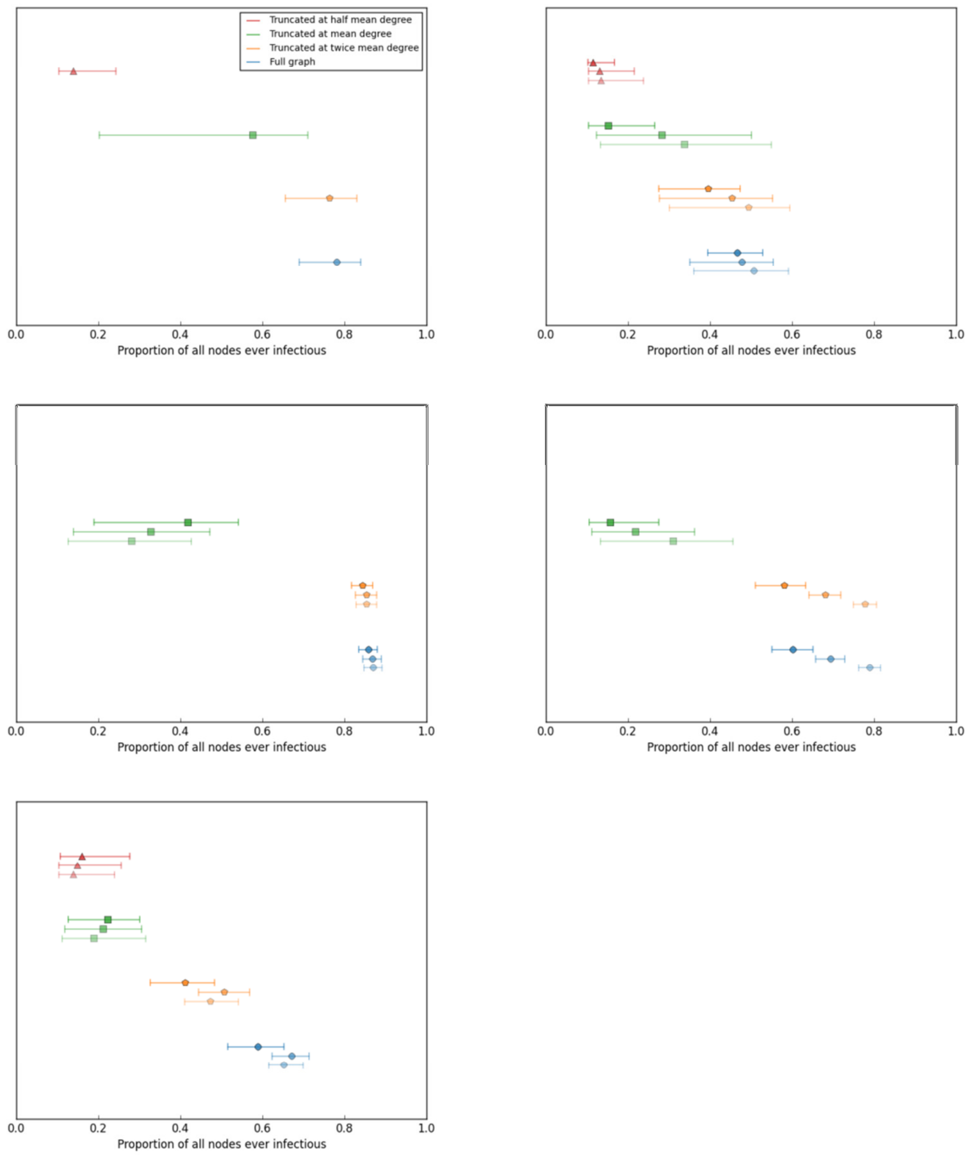
Figures are percentage points of 10,000 runs (synthetic networks) or 7500 runs (Karnataka villages).

Supplementary Figure 1: Time to infection of 10% of all individuals on networks, amongst epidemic simulation runs infecting at least 10% of the population



A: Karnataka villages; B: Degree-Assortative; C: Triadic Clustering; D: Focal Clustering; E: Power-Law networks. Figures show mean and 95% ranges for all runs from 10,000 simulations (7,500 for Karnataka villages) for which at least of 10% of individuals were ever infected. Simulation types are defined by truncation (see legend) and level of calibration – darker shading represents stronger calibration towards higher values of network properties (see **Error! Reference source not found.**). Empty lines represent simulation types where no runs reached the 10% threshold.

Supplementary Figure 2: Attack rate on networks, amongst epidemics infecting at least 10% of the population.



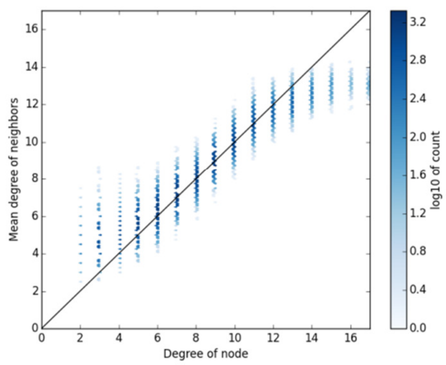
A: Karnataka villages; B: Degree-Assortative; C: Triadic Clustering; D: Focal Clustering; E: Power-Law networks. Figures show mean and 95% ranges for all runs from 10,000 simulations (7,500 for Karnataka villages) for which at least of 10% of individuals were ever infected. Simulation types are defined by truncation (see legend) and level of calibration – darker shading represents stronger calibration towards higher values of network properties (see **Error! Reference source not found.**). Empty lines represent simulation types where no runs reached the 10% threshold.

Supplementary Figure 3: Mean neighbor degree vs. own degree for full and truncated synthetic networks

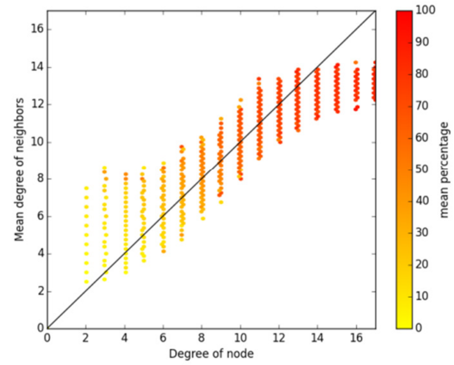
For each set of figures below:

A. Full graph; B: graph truncated at twice mean degree; C: graph truncated at mean degree; D: graph truncated at half mean degree. Within each cell, darker=more: Blue (A1): Initial density of ties (log-scale); Green (B1, C1, D1): Mean proportion of neighbors dropped (linear scale); Red-Yellow (A2, B2, C2, D2): Mean proportion of epidemic runs in which the node was infected (linear scale). The black diagonal line shows points of equal node and mean neighbor degree.

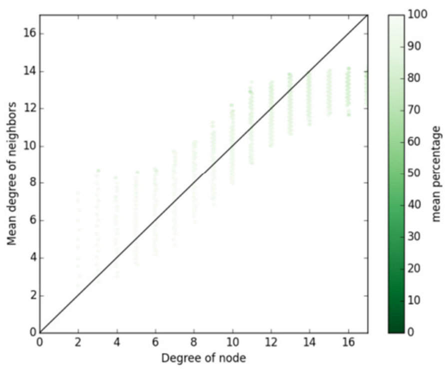
I. Degree-Assortative



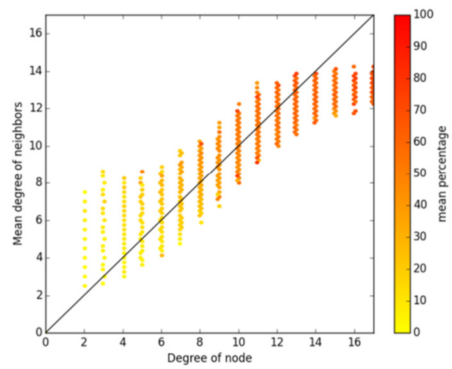
A1.



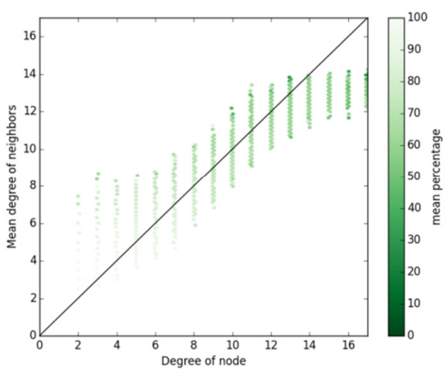
A2.



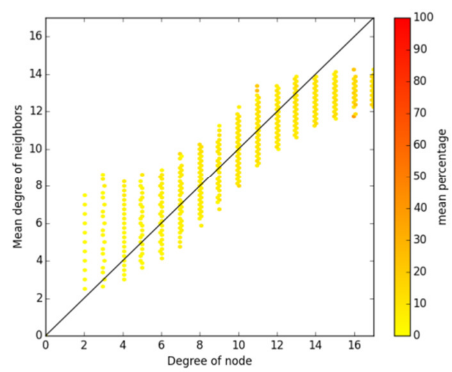
B1.



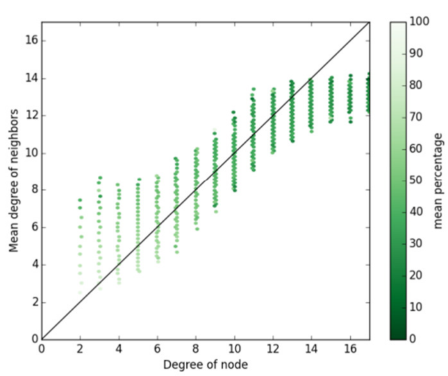
B2.



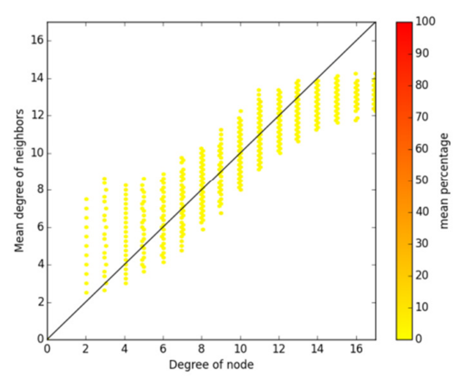
C1.



C2.

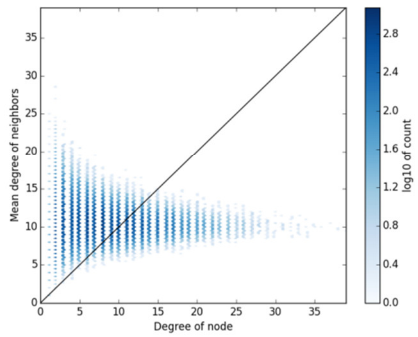


D1.

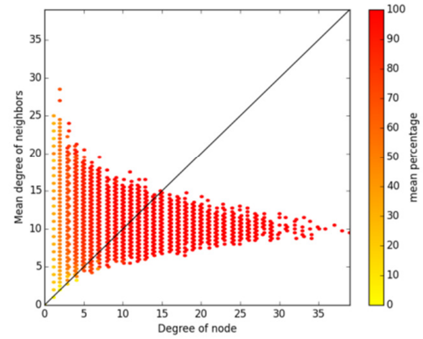


D2.

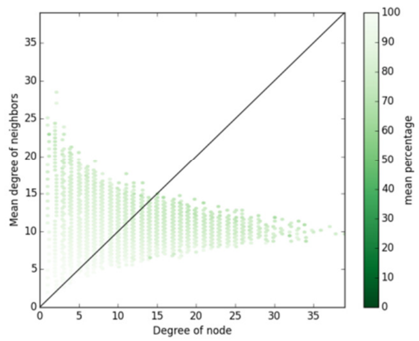
II. Triadic Clustering



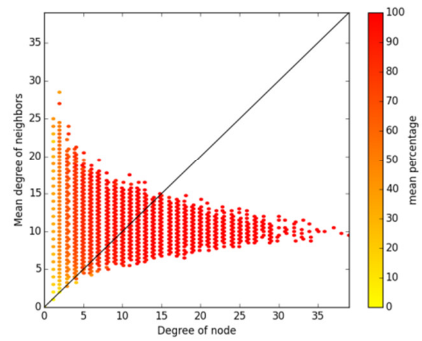
A1.



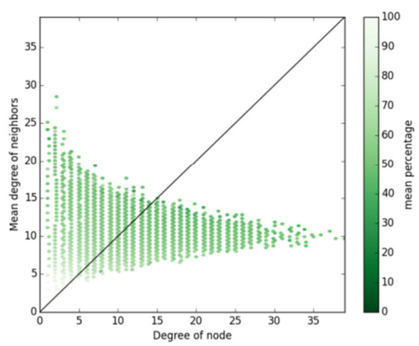
A2.



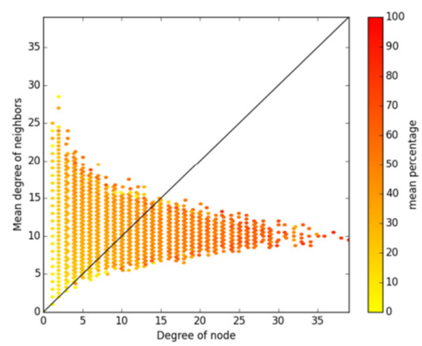
B1.



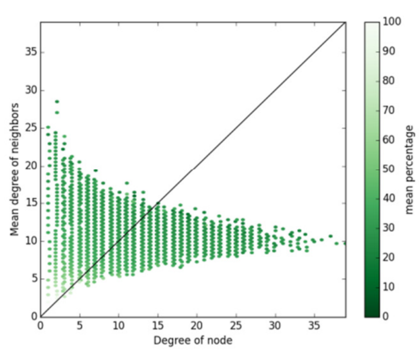
B2.



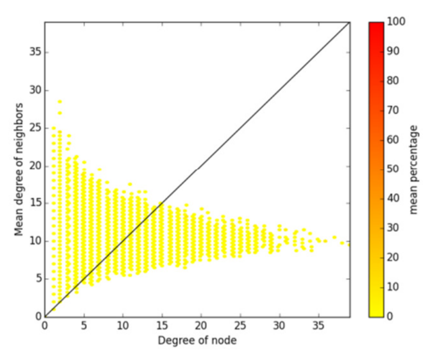
C1.



C2.

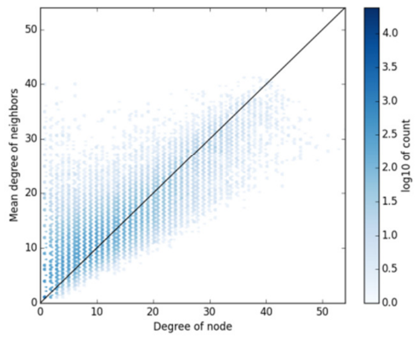


D1.

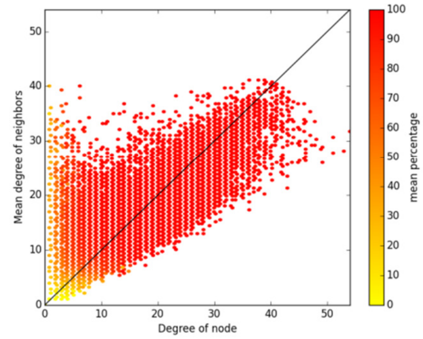


D2.

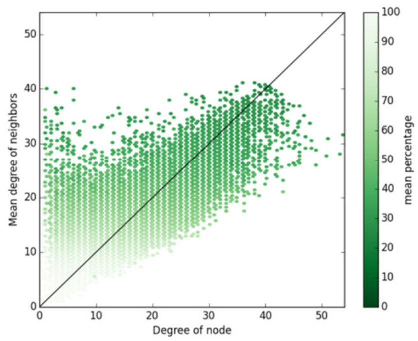
III. Focal Clustering



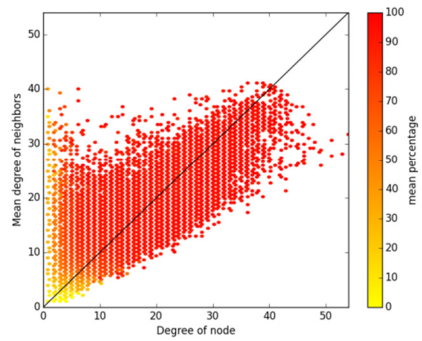
A1.



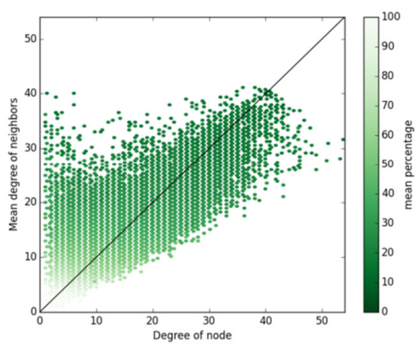
A2.



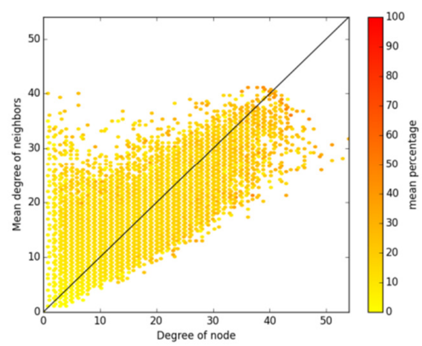
B1.



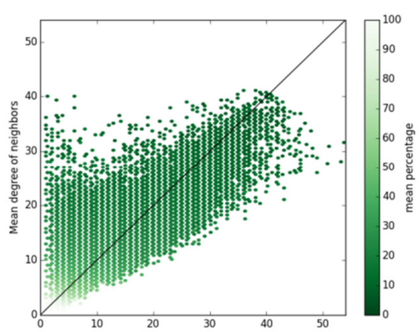
B2.



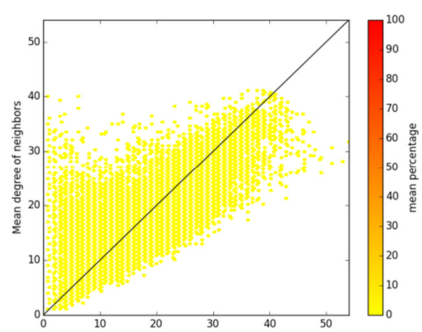
C1.



C2.

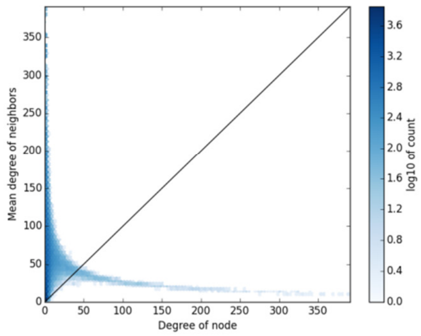


D1.

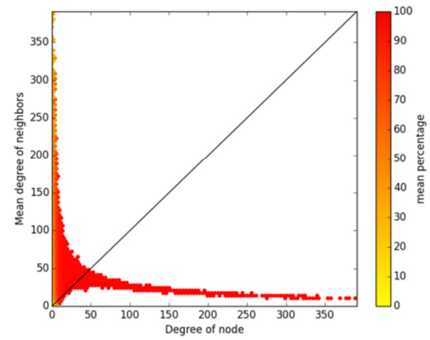


D2.

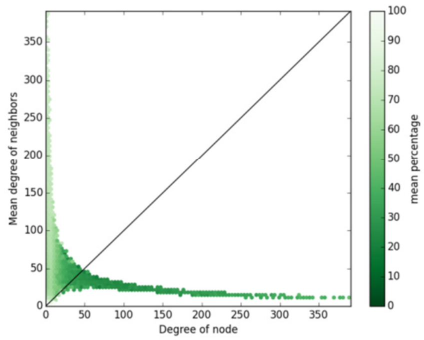
IV. Power-Law degree distribution



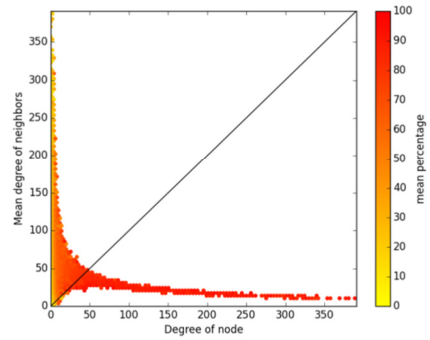
A1.



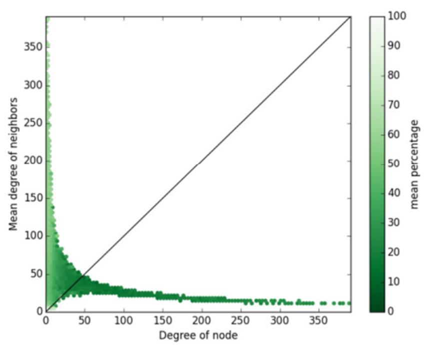
A2.



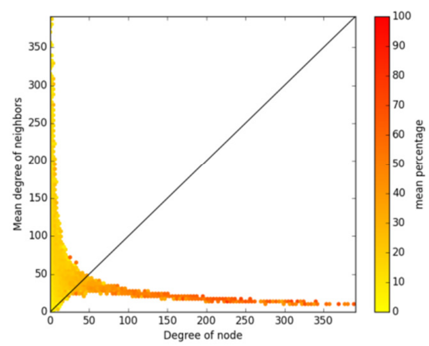
B1.



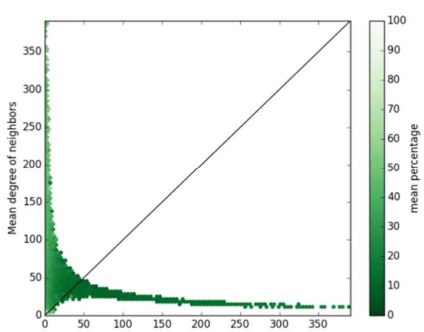
B2.



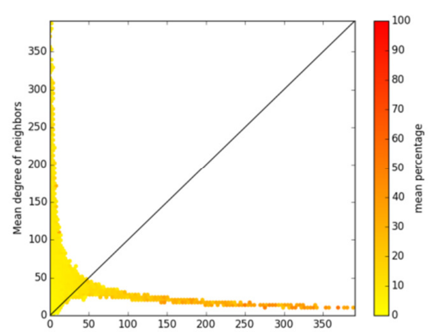
C1.



C2.

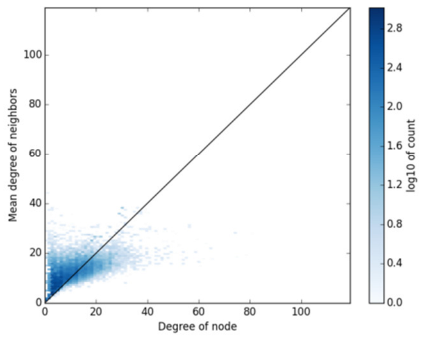


D1.

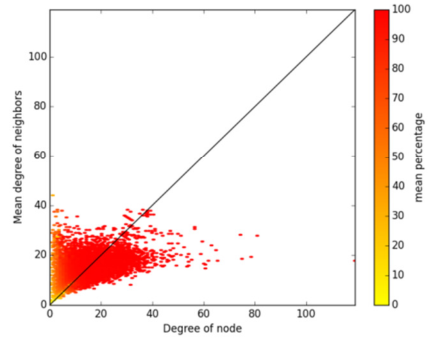


D2.

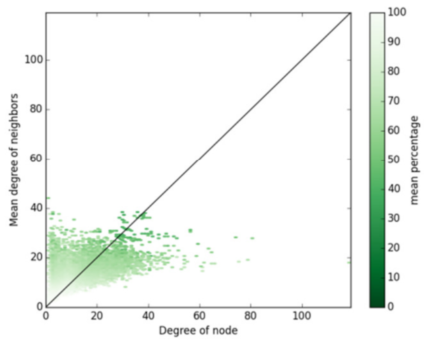
V. <u>Karnataka villages</u>



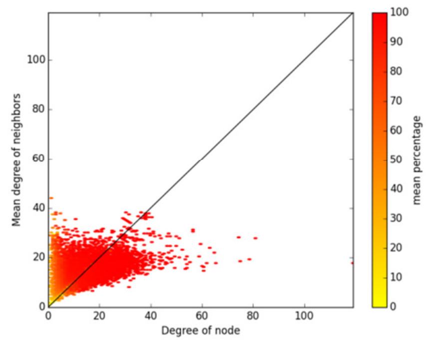
A1.



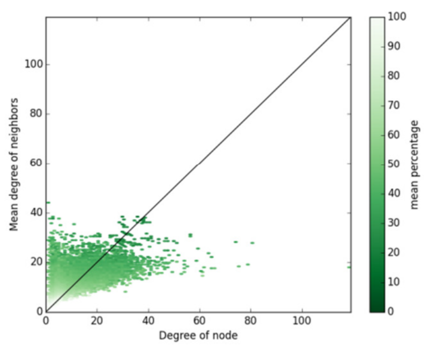
A2.



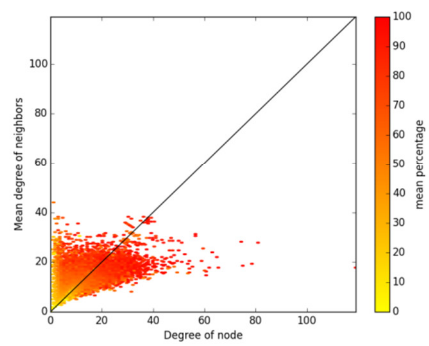
B1.



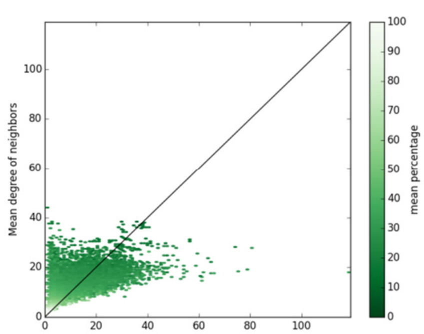
B2.



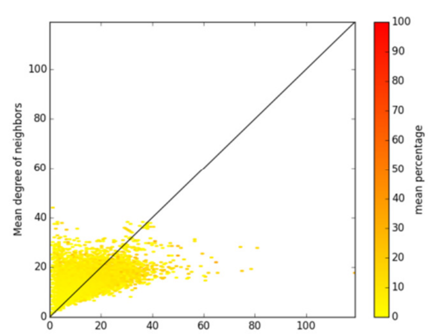
C1.



C2.



D1.



D2.