Supplementary Information for "Dynamic centrality measures for cattle trade networks"

1 Definition of Katz and betweenness centralities

Katz centrality. Whereas PageRank centrality measures the importance of a node by the time spent in the node by random walkers, the idea behind Katz centrality is to count the number of *paths* linking a given vertex to other vertices in the network. An attenuation parameter $\alpha \in (0, 1)$ is used to favor short paths over longer ones. The number of paths of length k between i and j is then the (i, j)-th element of A^k. Katz centrality is then defined by:

$$K(i) = \sum_{k=1}^{\infty} \sum_{j} \alpha^k (A^k)_{i,j}. \tag{1}$$

In a directed graph, the formula above remains valid, but only counts directed paths starting at i. A version of Katz centrality counting paths ending at i can of course be defined using the term $(A^k)_{j,i}$ in (1).

When α is sufficiently small (meaning $\alpha < 1/\rho(A)$, where $\rho(A)$ is the largest eigenvalue of A), the Katz centrality vector can also be computed using:

$$K = (I - \alpha A)^{-1} \mathbf{1}_n^{\mathsf{T}}, \tag{2}$$

where $\mathbf{1}_n = (1, \dots, 1)$.

Betweenness and closeness centrality. These two related centrality metrics rely on counting *shortest* paths passing through a given vertex. First, consider a vertex u in an undirected connected network $\mathcal{G} = (V, E)$. For each $v \in V$, let $d_{u,v}$ be the geodesic distance on V, *ie.* the length of a shortest path linking u and v. The closeness of u is then defined as

$$C(u)^{-1} = \sum_{v \neq u} d_{u,v}, \tag{3}$$

where the sum is in fact taken over all vertices v in the connected component of u. Intuitively, a vertex with high closeness can reach other nodes using short paths. Sometimes a normalized version of closeness centrality is considered, where each shortest-path distance is divided by the maximal length of a shortest path, namely $N-1$. In directed networks, one can consider separately an outgoing and an ingoing closeness centrality.

A similar idea lies at the heart of betweenness centrality: for three given vertices $u \neq v \neq w \in V$, let $\mathsf{sp}(u,w)$ be the number of shortest paths linking u to w and let $\mathsf{sp}(u,v,w) \leq \mathsf{sp}(u,w)$ be the number of those paths that pass through v. The betweenness centrality of v is then

$$B(v) = \sum_{\substack{u,w \in V \\ u \neq v \neq w}} \frac{\mathsf{sp}(u,v,w)}{\mathsf{sp}(u,w)}. \tag{4}$$

Exact computation of betweenness and closeness centrality can be challenging for large networks, since it relies on path enumeration and hence, to combinatorial explosion. Efficient methods have been developed [1] and implemented, for instance in the `networkit` package [2].

2 Variance of simulated outbreak final sizes

In Table 1, we present the mean and standard deviation of the final outbreak size of a temporal SIR outbreak started at a uniformly chosen node in the 2015 BDNI network. A fraction of nodes were removed according to their centrality computed using different methods. As we mentioned in the main text, standard deviations are quite low, with coefficients of variation uniformly below 5%, indicating that most outbreaks follow a similar pattern, probably driven by the existence of hubs.

Graph	Method	Mean outbreak size	Standard deviation
Complete	Betweenness	11,045	217
	Degree	11,524	221
	Katz	13,906	253
	PageRank	6,413	152
	out-TempoRank	286	15
	TempoRank	1,648	57
Reconstructed	Betweenness	24,633	367
	Degree	23,857	361
	Katz	23,231	359
	PageRank	24,247	364
	out-TempoRank	13,327	247
	TempoRank	2,566	82

Table 1: Mean and standard deviation of the final outbreak size in 10,000 simulated outbreaks on the 2015 BDNI network, with 0.025% of nodes removed according to decreasing centrality.

3 Fraction of markets and assembly centres in highly ranked nodes

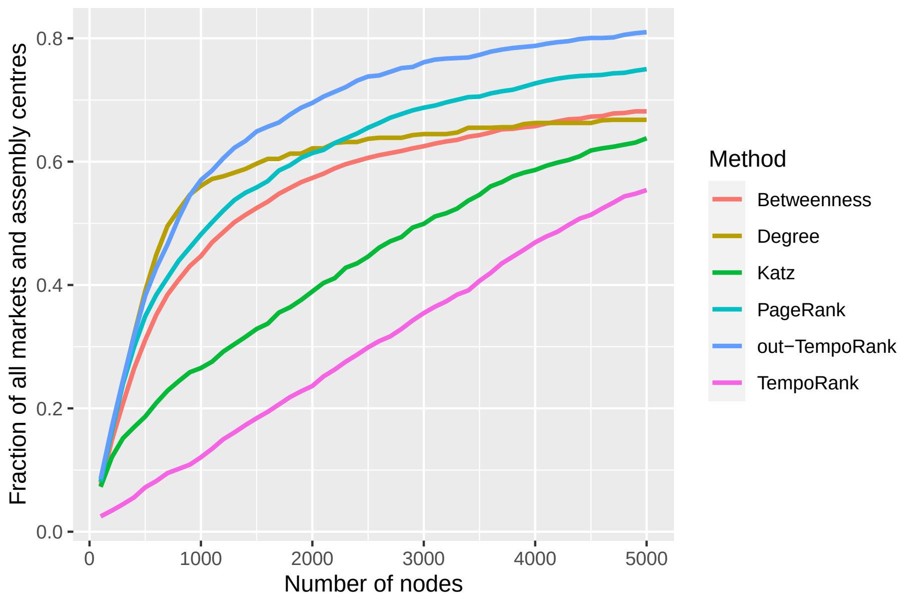

Figure 1: Proportion of markets and assembly centres in the highest-ranked nodes of the 2014 complete BDNI network.

In the complete BDNI network, there are 1,168 markets and assembly centres in the 2014 network (1,132 in the 2015 network). For $n \leq 5000$, we selected the n highest-ranked nodes according to the six centrality measures we considered in the paper, and computed the proportion of all markets and assembly centres that were present in those n nodes.

Four of the centrality measures considered (Betweenness centrality, degree centrality, PageRank centrality and out-TempoRank centrality) are highly ranking markets and assembly centres, with 50% of them consistently ranked in the 1,500 top-scoring nodes. The most effective centrality measure at identifying markets and assembly centres is out-TempoRank, followed by PageRank centrality. This hierarchy is reflected in the disintegration curves in the main text, in which PageRank centrality was the best-performing static measure.

Interestingly, as far as TempoRank centrality is concerned, there is a constant proportion of markets and assembly centres in the highest-ranked nodes. This proportion is lower than any of the other centrality measures, yet targeted removal manages to lower final outbreak sizes to a level that static centrality measures can not attain.

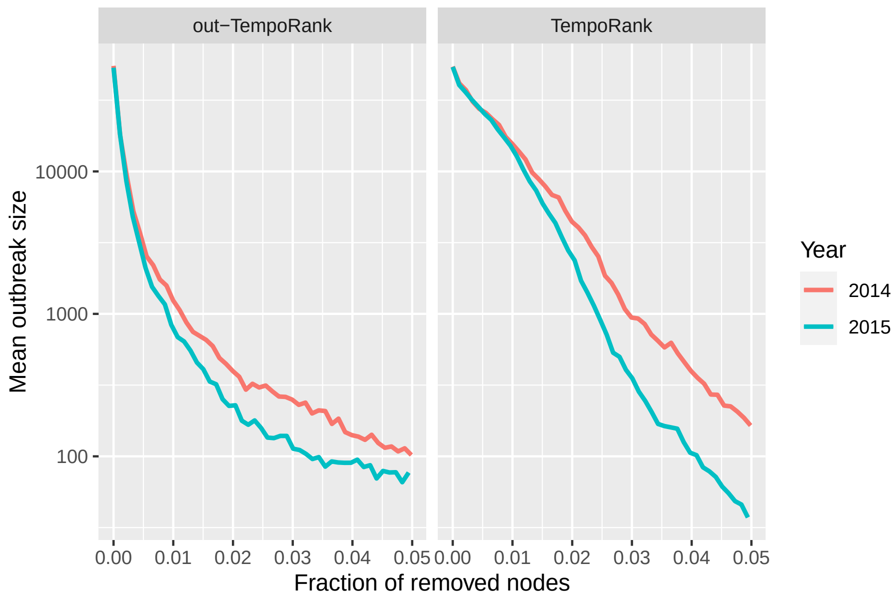

Figure 2: Mean outbreak size as a function of the fraction of removed nodes in the 2015 complete BDNI network, with centralities computed using the 2014 (red) or 2015 (blue) movement data (log scale).

4 Comparison of measures computed on different yearly networks

An important question is to ascertain the power of centrality measures computed retrospectively to predict epidemic outcomes simulated on contemporaneous data. As a benchmark, we compared the disintegration curves for centrality measures computed on the 2014 network with the curves for centralities computed on the 2015 network, all simulations being done using the 2015 movement data. Results are shown in Fig. 2 for the complete network and Fig. 3 for the reconstructed network.

Unsurprisingly, in both networks, the use of contemporaneous data improves the speed of decrease of mean final size, but the curves remain qualitatively similar, indicating that TempoRank and out-TempoRank scores are robust to the changes occurring at small scales on a year-to-year basis.

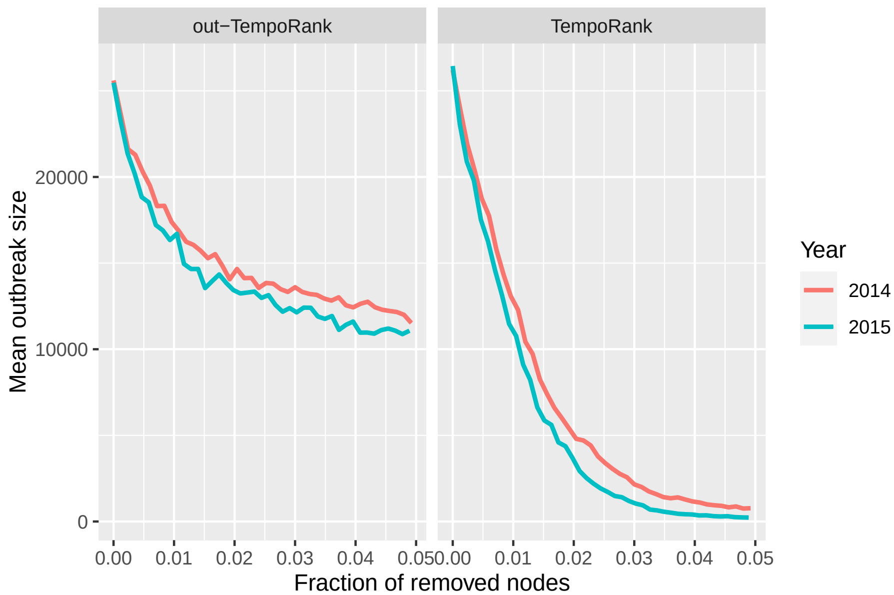

Figure 3: Mean outbreak size as a function of the fraction of removed nodes in the 2015 reconstructed BDNI network, with centralities computed using the 2014 (red) or 2015 (blue) movement data (linear scale).

5 Mean outbreak sizes as a function of epidemic parameters

In order to assess whether the hierarchy of centrality measures obtained through the percolation experiment described in the main text (Section 3.3) was robust to different spreading parameters, we removed a fixed amount of nodes (0.5% of active nodes in the 2014 network) from the 2015 complete BDNI network. We then simulated outbreaks for different values of the per-contact infection probability β and the per-node recovery rate γ. Results are shown in Fig. 4. In all cases, we recover the hierarchy found in the main text (corresponding to $\beta = 0.5$ and $\gamma = 1$).

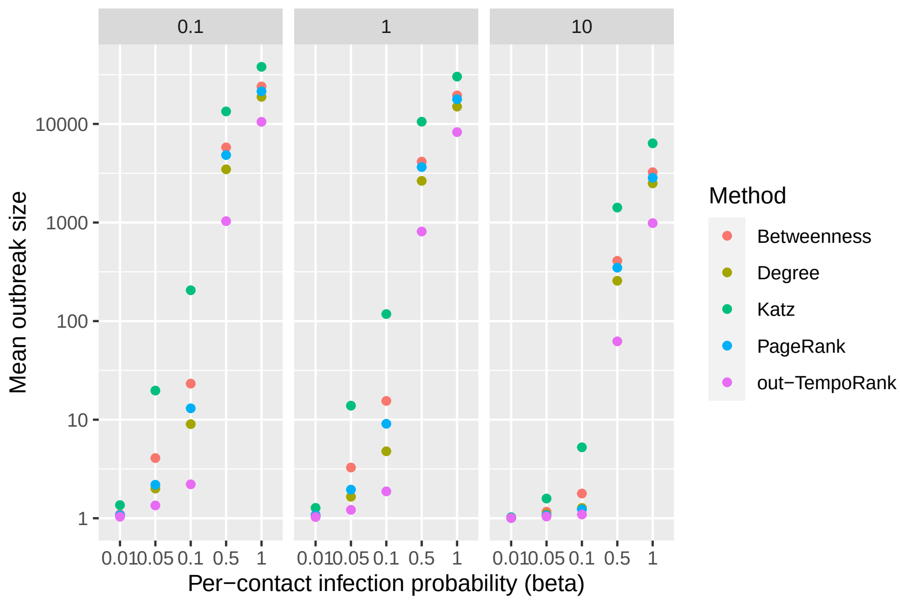

Figure 4: Mean outbreak size as a function of infection probability β. Facets correspond to the recovery rate γ.

6 Mean outbreak size as a function of random walk parameters

In Fig. 5 and 6, we present mean outbreak sizes, with 0.5% of nodes removed, for all studied combinations of parameters q and d, corresponding to the experiments described in Fig. 6 and 7 in the main text (which present the complete curves only for the slices $q = 0.1$ and $d = 0.15$). We recover the same phenomena already described in the main text, namely a strong effect for d, and a slight effect for q, which is more apparent for high d.

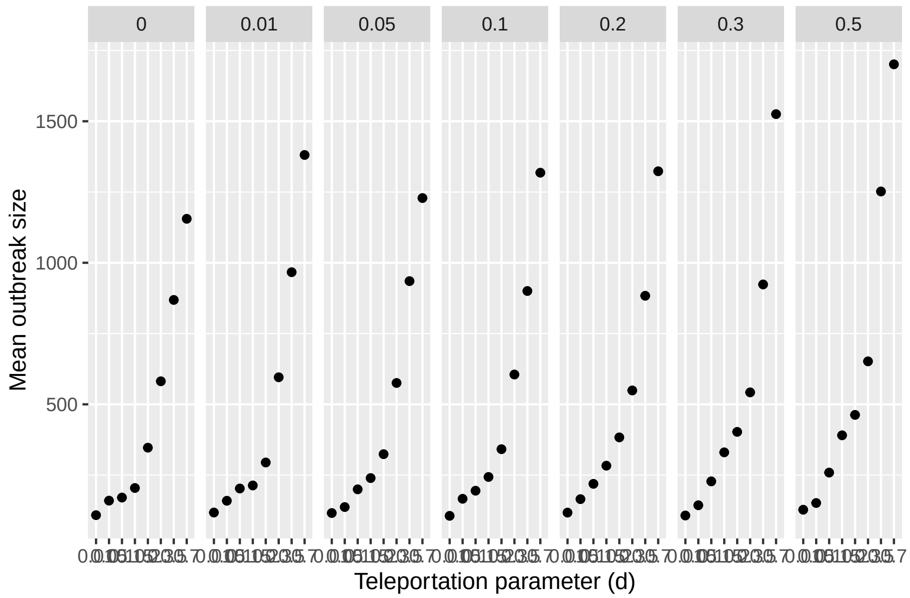

Figure 5: Mean outbreak size as a function of teleportation d (linear scale). Facets correspond to the laziness parameter q.

References

[1] Riondato, M., Kornaropoulos, E.M.: Fast approximation of betweenness centrality through sampling. Data Mining and Knowledge Discovery **30**(2), 438–475 (2016). doi:10.1007/s10618-015-0423-0

[2] Staudt, C.L., Sazonovs, A., Meyerhenke, H.: NetworKit: A tool suite for large-scale complex network analysis. Network Science **4**(4), 508–530 (2016). doi:10.1017/nws.2016.20. 1403.3005

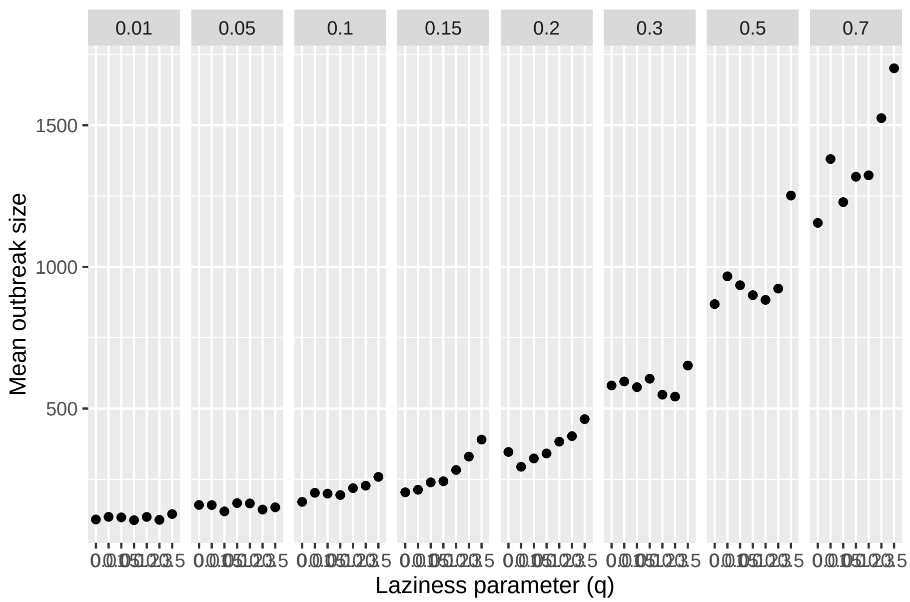

Figure 6: Mean outbreak size as a function of laziness q (linear scale). Facets correspond to the teleportation parameter d.