

## Structure and dynamics of diffusion networks

---

We are interested in developing machine learning and large-scale data mining methods for the analysis, modeling and control of large real-world networks and processes that take place over them. We addressed mainly two problems: network inference and influence maximization.

**Network inference** Observing a diffusion process often reduces to noting when nodes (people, blogs, etc.) reproduce a piece of information, get infected by a virus, or buy a product. Epidemiologists can observe when people become ill but they cannot tell who infected them or how many exposures and how much time was necessary for the infection to take hold. In information propagation, we observe when a blog mentions a piece of information. However if, as is often the case, the bloggers do not link to their sources, we do not know where they acquired the information or how long it took them to post it. Finally, viral marketers can track when customers buy products or subscribe to services, but typically cannot observe who influenced customers' decisions, how long they took to make up their minds, or when they passed recommendations on to other customers. We observe where and when but not how or why information (be it in the form of a virus, a meme, or a decision) propagates through a population of individuals. The mechanism underlying the process is hidden. However, the mechanism is of outstanding interest, since understanding diffusion is necessary for stopping infections, predicting meme propagation, or maximizing sales of a product.

In our work, we formulated a generative probabilistic model of diffusion that aims to describe realistically how diffusion occurs over time in a network. First, we developed two algorithms, NetInf [133, 566] and MultiTree [457], which use submodular optimization to infer the structure of a diffusion network from diffusion traces. We demonstrate the effectiveness of our algorithm by tracing information cascades in a set of 170 million blogs and news articles from 3.3

million sites over a one year period. We find that the diffusion network of news tends to have a core-periphery structure with a small set of core media sites that diffuse information to the rest of the Web.

However, both algorithms force the transmission rates between all nodes to be fixed and static – and not inferred. To overcome this limitation, we then developed the algorithm NetRate [69, 375, 418, 512], which allows transmission at different rates across different edges, possibly time-varying, so that we can infer temporally heterogeneous interactions within a network. NetRate infers both structure and the temporal dynamics of the underlying network with provable guarantees using convex optimization. We study the evolution of information pathways in the online media space and find that information pathways for general recurrent topics are more stable across time than for on-going news events. Clusters of news media sites and blogs often emerge and vanish in matter of days for on-going news events. Major events involving civil population, such as the Libyan's Civil War, lead to an increased amount of information pathways among blogs as well as in the overall increase in the network centrality of blogs and social media sites.

**Influence maximization** In this problem, one aims to select the most influential source node set of a given size in a diffusion network. A diffusion process that starts in such an influential set of nodes is expected to reach the greatest number of nodes in the network. Although the problem depends dramatically on the underlying temporal dynamics of the network, this was largely unexplored. In our work, we developed an influence estimation, InFluMax [458], which accounts for the temporal dynamics underlying diffusion processes [512]. Later, we developed a very efficient influence estimation method, ContInEst [429], which can be used in combination with InFluMax to scale up influence maximization to networks with million of nodes.

More information: <https://ei.is.tuebingen.mpg.de/project/structure-and-dynamics-of-diffusion-networks>