Network Structure

A Hyperbolic Space Analytics Framework for Big Network Data and Their Applications

E Stai, V Karyotis, S Papavassiliou - 《IEEE Network》, 2016, 30(1):11-17

Shuo Yu

2016-11-15
Big data analytics have generated a paradigm shift in modern data analysis and decision making in almost every aspect of human society. Nowadays, massive amounts of generated network and correlated (networked) data pose critical computational and storage challenges, requiring the development of radical techniques to manage, process, and analyze them more efficiently.
Big Data Analytics (BDA)
BDA have considerable impact on decision making, as well as on the machinery employed for collecting, processing, and transferring data. However, despite all the interest, effort, and resources dedicated, the anticipated BDA benefits have not come full circle yet. BDA seem to lie just in front of the dawn of the evolution to be witnessed. Especially with respect to big data derived from cloud applications and complex networks (big network data), or data exhibiting relations (correlations) among them, that is, structure that can be cast in network form (networked data), radically new approaches will be required to ensure scalability, efficiency, and longterm value of analytics compared to the currently employed approaches.
Workflow of the proposed HDA framework

**Network analysis and design**
- Link prediction
- Social metric approximation
- Greedy routing

**Resource allocation optimization**
- Advertisement allocation optimization
- Cross-layer design and optimization in multihop networks

**Network economics and marketing**
- Recommender systems
A demonstration of the process of discovering missing links based on hyperbolic distances is shown in red. In blue we illustrate an example of a path paved by greedy routing between points A and B, passing through the center of the hyperbolic space due to the tree structure. In orange we show the computation of social metrics (i.e., closeness centrality) based on hyperbolic coordinates. Given the coordinates of nodes C, D, and F, node C, which is embedded close to the center of the hyperbolic space, is expected to have lower average distance to all other nodes in the network compared to F. Indicatively, the hyperbolic distance between C and F \( d_H(z_C, z_F) \), where \( z_C = x_C + iy_C \), \( z_F = x_F + iy_F \) in the Poincare disk model is given by

\[
\cosh d_H(z_C, z_F) = \frac{2|z_C - z_F|^2}{1 - |z_C|^2(1 - |z_F|^2)} + 1.
\]
BDA typically involve large-scale optimization problems, which are expensive and suffer from slow numerical convergence rates, thus requiring new solution approaches. In this context, due to the scarcity of network resources, their allocation is one such crucial optimization problem that calls for efficient solution schemes. In this section, we study how HDA benefits large-scale (with respect to both the network and resources) resource allocation within two different frameworks: targeting (allocating) advertisements at the users of an online social network and cross-layer design/optimization in wireless decentralized networks. Note that the former problem can be cast in other resource allocation cases (e.g., file allocation/caching for fast retrieval in wireless heterogeneous networks).
HDA Applications

An example of an OSN’s users' allocation to five advertisers considering fairness with respect to the users' social influence (node degree). Each advertiser is assigned a pie-shaped area over the Poincare disk, on which the users’ OSN is embedded.
Delay performance of the Backpressure Algorithm (BP) vs. the source rate with and without HDA. BP is a module of the solution of a Network Utility Maximization (NUM) problem in wireless multihop networks. HDA significantly improves quality of service (QoS) (delay) values.
Assuming a complex network interconnecting customers and products.

**The first kind** aims to explore “how” one can recommend an item to a user (with particular characteristics/choices) with higher chances of acceptance, taking a sequential recommendation approach. The discovery of the most suitable recommendation chain is mapped to a greedy routing problem in hyperbolic space. Greedy routing can be intervened by choices of products from users that are similar to the target user, leveraging the correlations implied by the customers’ network (similar to collaborative filtering).

**The second type** of recommender system addresses the problem of recommending products to pairs of users with possibly non-identical interests. It applies greedy routing techniques combined with ant routing.
HDA Applications

Which chain of products should we recommend to user A for a smoother transition to recommending product $P_1$, increasing the possibility of its acceptance?

Assume that $P_A$ is a product that user A has chosen in the past (there is a vertical link between A and $P_A$). Then, applying greedy routing in hyperbolic coordinates from $P_A$ to $P_1$, a possible recommendation chain of products for user A is $P_2 \rightarrow P_3 \rightarrow P_4 \rightarrow P_5 \rightarrow P_1$.

Which product can one recommend to the pair of users A, B?

Suppose that $P_B$ is one of the products that user B has chosen in the past. We apply greedy routing from $P_B$ to $P_A$ and vice versa. The rendezvous point is the product $P_4$, which constitutes the output of the recommender system for the pair of users A, B. If the users have chosen more than one product in the past, all the pair combinations should be regarded for obtaining alternative recommendations from products lying on rendezvous points. This approach leverages on the close to perfect success of greedy routing in finding paths to the destination in hyperbolic geometry.
Network Structure

Structural and Collaborative Properties of Team Science Networks

Minh X. Hoang*, Ram Ramanathan†, Terrence J. Moore‡ and Ananthram Swami‡
Team Science Network
This is a simple collaboration network with three teams: $\{0, 1, 2\}$, $\{2, 3, 4\}$, and $\{1, 4, 5, 6\}$. The facet (team) sizes are 3, 3 and 4 respectively, and their dimensions are 2, 2, and 3 respectively. All subsets of the facets, including the facets themselves, are simplices and are shaded. The facet degree of vertex 4 (number of teams it belongs to) is 2 and that of vertex 5 is 1. Note that $\{1, 2, 4\}$ is not a simplex even though $\{1, 2\}$, $\{1, 4\}$ and $\{2, 4\}$ are simplices. Instead, $\{1, 2, 4\}$ is a 2-MNF.

Minimal non-face (MNF) may be used to identify “missed” collaborations. Vertices 1, 2 and 4 are all collaborating pairwise, which suggests a possible match in their interests. However, they miss an opportunity to combine their skills in a 3-way collaboration. A traditional graph-based model will fail to discern MNFs.
Network Structure

Network analytics in the age of big data

—-How can we holistically mine big data?

The four networks shown have exactly the same size (the same number of nodes and edges), and each node in each network has the same degree (the number of interactions with other nodes), but each network has a very different structure.
Network Structure
A Three-layer Vector Space Network Structure
—How to describe Scientific Team Science network?
A Three-Layer Network

- Co-author Network
- Institution Collaboration Network
- Citation Network
- Other Networks
A Three-Layer Vector Space Network
Thank you