PhD thesis proposal on

Communities in Networks: Detection of dense subgraphs and partitions

Supervisor

Cristina Bazgan, LAMSADE, Université Paris-Dauphine, cristina.bazgan@dauphine.fr

Description

Communities can be formed by regrouping elements (people, web sites, ...) according to some relation between these elements (based on affinity, common interest, ...). The detection of communities or partitions into communities in networks (the world wide web, communication networks, social networks, biological networks) has many natural applications [8] including, e.g., designing efficient recommendation systems or identifying clusters of customers. These networks are modeled by a graph where arcs/edges represent relationships or interactions within a group of individuals. Intuitively a community is a "dense" subgraph. Depending on the context, there are different ways of defining communities or dense subgraphs.

The most natural way to define a dense subgraph is to consider the complete graph. Several other definitions were proposed in the literature like k-core (every vertex has degree at least k in the subset), k-clubs (the subset has a diameter at most k), or α-quasicliques (the subset contains at least a ratio α of the maximum number of edges). Another way of defining the density of a subset is the average degree of the subset. For the latter definition, finding a densest subgraph is polynomial-time solvable [9], finding a densest subgraph with lower or upper bounds on the cardinality is NP-hard [10], and finding a partition with a maximum sum of densities is NP-hard and approximable [5]. In [4], it is proved that the problem of finding a dense subgraph that is well-connected in terms of vertex/edge connectivity is approximable. In this context finding a dense subgraph that is also a k-club or finding a partition in such subgraphs would be interesting. In the context of social networks, Olsen [11] defines a community as a vertex set such that each vertex is proportionally "more related" to its own part as to the others. Such communities are computable in polynomial time. In the web context, a community is defined in [6] as a set of web pages that links to more web pages in the community than to pages out of community. Interesting weighted extensions are studied in [7]. The problem of partitioning into communities amounts to determining a satisfactory k-partition, for a given or suitably chosen k [3]. In [1, 2], a vertex of a subgraph is happy if it has all its neighbors inside the subsets. Several interesting questions remain to study about the complexity of finding a subset or a partition into subsets with a maximum number of happy vertices.

Objectives

The goal of this PhD is to identify interesting definitions of density that appear in applications and to define algorithmic approaches based on existing or new such definitions. Generally, most of these optimization problems are NP-hard. The goal is to identify cases for which the problems are polynomial-time solvable and establish exact and approximation algorithms for solving these problems. Other approaches like exact and approximate parameterized algorithms could also be useful.

Candidate profile

The ideal candidate has a keen interest and strong background in combinatorial optimization, graph theory, exact complexity. Notions on parameterized complexity and approximation algorithms are a plus. She/he has to have a Master degree or equivalent in these areas. He/she should have excellent academic qualifications.

Position

The position is fully funded for 3 years. There is no obligation to teach, but it will be possible (with additional payment). Funding for attending international conferences, summer schools, and visiting other research centers (Trier University, TU Berlin, LAU in Beyrouth, Portsmouth University) will also be provided. Starting date is expected in September-October 2022.

Application

Any interested candidate is invited to email his/her application (including CV, letter of motivation, academic record during the last two years and possibly the name and email of reference persons) to cristina.bazgan@dauphine.fr by May 1st, 2022.

References

[1] Y. Asahiro, H. Eto, T. Hanaka, G. Lin, E. Miyano, and I. Terabaru, Parameterized algorithms for the Happy Set problem, *Discrete Applied Mathematics*, 304, 32-44, 2021.

[2] Y. Asahiro, H. Eto, T. Hanaka, G. Lin, E. Miyano, and I. Terabaru, Complexity and approximability of the happy set problem, *Theoretical Computer Science*, 866, pp. 123-144, 2021.

[3] C. Bazgan, Zs. Tuza, and D. Vanderpooten, The satisfactory partition problem. *Discrete Applied Mathematics*, pp. 1236-1245, 2010.

[4] F. Bonchi, D. García-Soriano, A. Miyauchi and C. E. Tsourakakis, Finding densest k-connected subgraphs, *Discrete Applied Mathematics*, 305, 34-47, 2021.

[5] J. Darlay, N. Brauner, and J. Moncel, Dense and sparse graph partition, *Discrete Applied Mathematics*, 160, 16-17, pp. 2389–2396, 2012.

[6] G. Flake, S. Lawrence, and C. Lee Giles, Efficient identification of web communities. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 150-160, 2000.

[7] G. Flake, R. Tarjan and K. Tsioutsiouliklis, Graph clustering and minimum cut trees. *Internet Mathematics*, 1(3), 355-378, 2004.

[8] S. Fortunato, Community detection in graphs. *Physics Reports*, 486, 75-174, 2010.

[9] A. V. Goldberg, Finding a maximum density subgraph. *Technical report*, 1984.

[10] S. Khuller and B. Saha, On Finding Dense Subgraphs. *ICALP 2009*, LNCS 5555, pp. 597-608.

[11] M. Olsen, On defining and computing communities. *ACSC 2013*, pp. 23-30.