Book of Poster Abstracts

A one-day workshop on

Computational Aspects of Complex Networks

Rome, December 6, 2024

Scientific and Organizing Committee:

- Daniele Bertaccini (University of Rome Tor Vergata, Italy)
- Andrea Clementi (University of Rome Tor Vergata, Italy)
- Carla Manni (University of Rome Tor Vergata, Italy)
- Mariarosa Mazza (University of Rome Tor Vergata, Italy)

The workshop is part of a series of scientific activities of the MUR Excellence Department Project MatMod@TOV (CUP E83C23000330006)

Disassortative dynamic BA models inspired by the Bitcoin Lightning Network

Mohamed Taki Eddine Abedesselam, University of Rome "Tor Vergata"

The evolution of the channel graph of the Lightning Network, the main layer-2 solution on top of Bitcoin, has been often modeled as a Barabási-Albert random graph. However, the available data on the channel graph of the Lightning Network indicate that it is stabilizing over a network structure with negative assortativity, while the assortativity of a BA random graph tends to zero as the number of nodes grows to infinity. In the BA model, new edges are created when a new node joins the network and they never disappear. In the channel graph of the Lightning Network, channels can be closed by one or both of its endpoints at any time. In this paper, we propose two dynamic versions of the BA-model in which edges can disappear at any time. In the first version, the edge disappearance rate is a function of the degrees of the endpoints, in the second one it is a function of the edge capacity. Simulations show that, in both models, the assortativity converges to a negative value, that depends on the edge disappearance rate. Our results suggest that the disassortative nature of the Lightning Network, as well as that of several other real networks, could be a consequence of the dynamic nature of the edges of the network.

The poster is based on joint work with Fabio Giacomelli and Francesco Pasquale (University of Rome "Tor Vergata").

Statistical validation of weighted projections of bipartite networks

Lorenzo Buffa, University of Rome "Tor Vergata" and Enrico Fermi Research Center, Rome
Daniele Cirulli, University of Rome "Tor Vergata" and Enrico Fermi Research Center, Rome

Bipartite graphs capture two-mode relationships between two disjoint groups of nodes and are widely used to model complex systems in economics, ecology, and social sciences. Within these contexts, indirect relationships among nodes in the same group are typically obtained by projecting the bipartite network onto one of its layers. However, the resulting one-mode network is typically very dense, as any two nodes are connected as soon as they share one neighbor in the other set, hindering the identification of relevant connections.
In this work we introduce a novel method for analyzing projected one-mode networks derived from weighted bipartite graphs, by inferring the statistical significance of each link in the projected network, thus effectively filtering meaningful connections out of those introduced by structural bias and random noise.
Our statistical validation algorithm can be summarized as follows:

1. Construct the empirical weighted bipartite network from data and compute the weighted projection on the selected layer, by compressing the adjacency matrix on the other layer.

2. Build a null model through a maximum entropy procedure, which constructs an ensemble of networks that are maximally random but preserve some properties of the empirical network. Model parameters are obtained by iteratively solving a system of N coupled non-linear equations, where N is the number of nodes. Determine the null model's distribution by sampling network realizations from the ensemble.

3. Assess the relevance of each link in this projection by comparing its empirical value against its null model distribution. Compute the corresponding p-value and statistically validate the link if the p-value is below a chosen threshold (determined through multiple hypothesis testing procedures such as FDR, Bonferroni, etc.).

In this way our method singles out the connections that are unlikely realizations of the null model, thereby producing a refined one-mode network that retains high-confidence relationships while reducing noise.
The selection of the null model in step 2 deeply influences the final result of the validation. In our work we introduce a model that constrains both the binary and weighted degree of each node in the projection, and compare it with the main methods currently available in the literature (that can be seen as an approximation of our full null). Application of our method on real datasets from economic, ecological, and social domains, as well as synthetic data (Erdos-Renyi, Block model, etc.), demonstrate that our method enhances interpretability and mitigates the inclusion of biased connections in the projection. Our approach thus represents an effective tool for researchers and practitioners working with complex bipartite networks, enhancing the understanding of network structures and indirect relationships within them.

The poster is based on joint work with Giulio Cimini (University of Rome "Tor Vergata"), Dario Mazzilli, Aurelio Patelli, Fabio Saracco (Enrico Fermi Research Center, Rome).

Katz-like centrality algorithms in Alzheimer and cancer diseases' detection

Luigi Chiricosta, University of Rome "Tor Vergata"

Biological systems, such as cells, are the results of huge amount of molecular components that interact each other at different level. The level of expression of transcripts, proteins and metabolites are also able to influence the way in which the interactions take place giving to the cell a dynamic behaviour. Bioinformatics databases such as STRING and KEGG were built to collect the most of the information about the static interactions of the cell molecules. In detail, STRING collects the knowledge about protein-protein interactions from different sources. The source can be experimentally validated or inferred by literature. In this sense, a confidence score is associated to each interaction. The interactions included in STRING are mainly undirected and are related to the protein itself. Nevertheless, they do not take in consideration neither the cell status nor the physiological or pathological condition. Actually, the environment is quite important in the study of the biological interactions. The cell system itself can be subject to a modification in its behaviour during time or after external stimuli. In this sense, KEGG collects the pathways information of the cell system. Pathways represent directed interactions that are specific in a particular biological context or during a disease.
In bioinformatics, the analysis of the protein-protein interactions is mainly based the study of the network through number of nodes, number of edges, average number of neighbors, network diameter, network radius, characteristic path length, clustering coefficient, network density, network heterogeneity, network centralization, connected components. Additionally, the typical scores that are used to identify cardinal proteins of the network are average shortest path length, clustering coefficient, closeness centrality, eccentricity, stress, degree, betweenness centrality, neighborhood connectivity, number of edges, radiality, topological coefficient. Nevertheless, most of these scores cannot be computed efficiently, a severe issue when these are large networks. Recently, other scores that can be computed much more efficiently were proposed such as Katz centrality, Eigenvector centrality or subgraph centrality. Unfortunately, even if each of them can provide useful information of the network, they do not take in consideration the biological context and the dynamical behaviour of the cell system.
Here we propose a new centrality index that integrates the biological context of the cell system. The new score is based on a Katz-like centrality algorithm. With this we plan to give relevance to source-to-sink walks of the underlying pathways. The biological reason under this is based on the fact that the cell signal is often triggered by external stimuli (source) that trigger, in cascade, other molecules that, in the end, lead a specific biological process (sink). We compare this new index against the classical used scores in specific real biological contexts such as under diseases (Alzheimer's Disease, Cancer Disease), biological process (Apoptosis, Cell cycle) and environmental information processing (MAPK signaling, PI3K-AKT signaling).

The poster is based on a joint work with Daniele Bertaccini and Alessandro Filippo (University of Rome "Tor Vergata").

Mittag-Leffler Communicability Distance and shortest geodesics on networks

Nikita Deniskin, Scuola Normale Superiore, Pisa

The Mittag-Leffler functions interpolate between the exponential and the resolvent functions, and they can be used for walk-based indices in networks, allowing to tune the weight of longer closed walks. Here we study the Communicability based on the Mittag-Leffler function, and define Mittag-Leffler communicability distance functions, which are proved to be Euclidean and spherical. We then geometrize the graphs based on these distances and obtain the corresponding geodesics based on them. We finally analyze these geodesics for different kinds of real-world networks.

The poster is based on a joint work with Ernesto Estrada (Institute of Interdisciplinary Physics and Complex Systems, Palma de Mallorca).

Role extraction by random walks and matrix equations

Dario Fasino, University of Udine

The nodes in a network can be grouped into equivalence classes according to the role they play. The grouping is based on their connections with nodes in either the same role or different roles. Browet and Van Dooren introduced a node similarity matrix to solve this role extraction problem in directed graphs, which is defined in terms of the solution of a suitable matrix equation. In a later step, the similarity matrix allows to group nodes with the same role. However, this construction tacitly relies on the assumption that the connections within and between the different roles are fairly uniform, a condition that is rarely met in real-world networks. Here, I propose a variant of the Browet-Van Dooren method that relies on an appropriate diagonal scaling of the adjacency matrix to compensate for the inhomogeneity of node connections. Compared to the Browet-Van Dooren method, theoretical properties and numerical experiments show better performance.

The work was carried out as part of the PRIN22 project entitled 'Low-rank Structures and Numerical Methods in Matrix and Tensor Computations and their Application - 20227PCCKZ'.

New numerical strategies for approximately updating Katz centrality

Alessandro Filippo, University of Rome "Tor Vergata"

Katz centrality is one of the most popular indices used in Network Science to determine the relative importance or influence of a node within an undirected network. A unique feature of Katz centrality, however, is that it can be rewritten as the solution of a symmetric linear system, meaning that it is computable by using broadly available linear algebra methods, like the conjugate gradient. Nevertheless, repeated calculation of the measure can still be challenging for large evolving networks when several linear systems need to be solved to maintain the scores up-to-date.
Here, we analyze how low-rank changes in the adjacency matrix, such as node or edge deletions, impact the centrality scores of the remaining nodes and propose efficient strategies to approximate updating Katz centrality, reducing the need for full linear system recalculations.

The poster is based on a joint work with Francesca Arrigo (Strathclyde University) and Daniele Bertaccini (University of Rome "Tor Vergata").

Computation of temporal centralities in regional water networks

Francesco Gravili, Alma Mater Studiorum-Università di Bologna

Multilayer networks are often used to model systems where entities are connected in multiple ways simultaneously, capturing the complexity of real-world relationships better than traditional single-layer networks. The dynamical evolution of a complex system over time can be described through a particular interlayer structure. The key idea of our project stems from the application of complex network theory to the analysis of a zonal water system, in this case the water distribution network of the Emilia-Romagna region. The ability of water networked systems to continue to supply water when components are damaged or fail is a central topic in the water system design and analysis. This feature of water systems is related to the connectivity of the internal components of the system (pipes and reserves). In this particular case, we assumed the network to be undirected, which implies its adjacency matrix to be symmetric. One important feature of our data set is a detailed account of all the pipe breakages in the water system over the course of time, from a few hours up to many months. To appropriately track the phenomenon, we propose to model breakages as low-rank modifications of the edge adjacency matrix of a water network within a time-aware block influence matrix. Our model is designed to better account for changes in centrality of different pipes over time, monitoring the influence of the number of breakage events occurring in time. Multilayer networks typically collect information of all time instances into a single large matrix, exposing all adjacency matrices. For this reason, computation of centralities of temporal multilayer networks can be very challenging, requiring specifically designed numerical procedures to make this computation feasible and reliable. Our model aims to meet these goals, while providing insightful results on a real pipeline network.

Propagate: a seed propagation framework to compute Distance-based metrics on very large graphs

Antonio Cruciani, Gran Sasso Science Institute, L'Aquila
Daniele Pasquini, University of Rome "Tor Vergata"

We propose Propagate, a fast approximation framework to estimate distance-based metrics on very large graphs such as: the (effective) diameter or the average distance within a small error. The framework assigns seeds to nodes and propagates them in a BFS-like fashion, computing the neighbors set until we obtain either the whole vertex set (for computing the diameter) or a given percentage of vertices (for the effective diameter). At each iteration, we derive compressed Boolean representations of the neighborhood sets discovered so far. The Propagate framework yields two algorithms: Propagate-P, which propagates all the m seeds in parallel, and Propagate-S which propagates the seeds sequentially. For each node, the compressed representation of the Propagate-P algorithm requires m bits while Propagate-S 1 bit only. Both algorithms compute the average distance, the effective diameter, the diameter, and the connectivity rate (a measure of the the sparseness degree of the transitive closure graph) within a small error with high probability. The time complexity of our approaches is $O(D \cdot m)$ for Propagate-P and $O(\log n/eps^2 D \cdot m)$ for Propagate-S, where m is the number of edges of the graph and D is the diameter. The experimental results show that the Propagate framework improves the current state of the art in accuracy, speed, and space.

The poster is based on a joint work with Giambattista Amati, Simone Angelini (Fondazione Ugo Bordoni, Rome), and Paola Vocca (University of Rome "Tor Vergata").

Explicit characterization of consensus in a distributed estimation problem on chain graphs

Cristiano Maria Verrelli, University of Rome "Tor Vergata"

This Abstract briefly describes the original results that are to be submitted to the international journal IEEE Transactions on Automatic Control. In particular, the problem of looking for time-varying vectors $\hat{\Theta}(t)$ - named Θ-estimates - that exponentially converge to the unknown constant parameter vector $\Theta \in \mathbb{R}^m$ defined by the set of linear time-varying equations:

$$y_1(t) = \phi_1^{\mathrm{T}}(t)\Theta, \tag{1}$$

$$y_i(t) = \phi_i^{\mathrm{T}}(t)\Theta, \quad i = 2, \ldots, p-1 \tag{2}$$

$$y_p(t) = \phi_p^{\mathrm{T}}(t)\Theta, \tag{3}$$

is addressed, where y_i are the locally measured outputs and $\phi_i(\cdot) : \mathbb{R}_0^+ \to \mathbb{R}^m$ are the local regressor vectors, $i = 1, \ldots, p$, each of them assumed to be available at the running time at each node of the graph. This general problem is referred to as the estimation problem for (1)-(3) on a graph. Now, when there are no edges connecting nodes of the graph that are far from each other, the problem passes from being full-graph-knowledge based to being partial-graph-knowledge based. This way, the burden of information that has to be communicated to the various measurement/estimation nodes of the graph might be (even largely) reduced. Specifically, the original contribution consists in showing that, under the weakest Θ-identification condition on the graph [2], a set of suitably tailored differential equations for the time-varying vectors $\hat{\Theta}^{[i]}(t)$, all of them converging to the unknown Θ and by thus achieving consensus, can be (redundantly) designed at each node $i = 1, \ldots, p$. This is proved under the condition that nodes undirectedly connected in series (undirected chain graph [3]) are considered, so that each estimation scheme at the node can share information - namely, its own Θ-estimate - with the neighbours only, one for node 1 and p, two for the remaining ones. Thus the aforementioned problem belongs to the class of problems known as multi-agent estimation with consensus. The general scenario is the one given by a multisensor network in which a distributed parameter estimator is to be designed on the basis of distributed sensing. This situation typically arises in the case in which mobile sensors measure the distribution of an unknown quantity over a field and, due to the size of the region, it is computationally inefficient or just infeasible to visit every point in the space to collect data ([4, 1, 7, 5]). Now, agents at the nodes face a local identification problem, in which they cannot consistently estimate the parameter vector in isolation so they have to engage in communication with their neighbours. In particular, estimate-consensus has to be achieved through a sort of penalization of the mismatch between the parameter estimates. Indeed, the original results of this Abstract regard the positive definite nature of a quadratic form associated with a tridiagonal block structure. Such a quadratic form is crucial to innovatively prove the exponential achievement - with an explicit characterization of the exponential convergence - of the distributed parameter estimation task on an undirected chain graph, in which an original neighbourhood-based decentralized parallel architecture (aiming at reducing the computational burden of a centralized estimation scheme on the graph) is adopted. It is not necessary for Θ to be fully identifiable at each node. Instead, it is sufficient for the nodes in the graph to cooperatively provide information to the local adaptive laws. Indeed, the derivations here move along the direction of [2] and are in exact accordance with it. However, in contrast to [2], which - for general graph topologies - uses weaker contradiction arguments to

prove exponential consensus, here original proofs of convergence are able to provide an explicit characterization of the exponentially achieved consensus. Nevertheless, the problem of identifying time-varying parameters that are periodic with known periods can be innovatively solved as well. It constitutes an additional original contribution of the paper: adaptive tools can be directly replaced by repetitive learning tools [6] within the same theoretical framework, where the asymptotic consensus is successfully ensured under identification mechanisms relying on the crucial role of the information exchange between neighbours.

The poster is based on a joint work with Carmine Di Fiore and Patrizio Tomei (University of Rome "Tor Vergata").

References

[1] H. Adjerid and A.R. Maouche, Multi-Agent system-based decentralized state estimation method for active distribution networks, Computers & Electrical Engineering, vol. 86, 2020.

[2] W. Chen, C. Wen, S. Hua, and C. Sun, Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus, IEEE Transactions on Automatic Control, vol. 59(1), pp. 91-106, 2014.

[3] A. Concas, L. Reichel, G. Rodriguez, and Y. Zhang, Chained graphs and some applications, Applied Network Science, vol. 6(39), 2021.

[4] Q. Liu, S. Zheng, Z. Wang, N.D. Alotaibi, and F.E. Alsaadi, Distributed estimation for multi-agent systems with relative measurements and quantized communication: A feedback quadratic framework, International Journal of Robust and Nonlinear Control, vol. 33(5), pp. 3164-3184, 2023.

[5] A.S. Matveev, M. Almodarresi, R. Ortega, A. Pyrkin, and S. Xie, Diffusion-based distributed parameter estimation through directed graphs with switching topology: Application of dynamic regressor extension and mixing, IEEE Transactions on Automatic Control, vol. 67(8), pp. 4256-4263, 2022.

[6] C.M. Verrelli, Exponential output tracking via a saturated RLC for a class of nonlinear systems under PSD of periodic references, Systems and Control Letters, vol. 192, 2024.

[7] J. Yan and H. Ishii, Distributed parameter estimation under Gaussian observation noises, IFAC-PapersOnLine, vol. 56(2), pp. 3960-3965, 2023.

Maintaining k-MinHash signatures over fully-dynamic data streams with recovery

Luca Pepè Sciarria, University of Rome "Tor Vergata"
Alessandro Straziota, University of Rome "Tor Vergata"

We consider the task of performing Jaccard similarity queries over a large collection of items that are dynamically updated according to a streaming input model. An item here is a subset of a large universe U of elements. A well-studied approach to address this important problem in data mining is to design fast-similarity data sketches. In this paper, we focus on global solutions for this problem, i.e., a single data structure which is able to answer both Similarity Estimation and All-Candidate Pairs queries, while also dynamically managing an arbitrary, online sequence of element insertions and deletions received in input.
We introduce and provide an in-depth analysis of a dynamic, buffered version of the well-known k-MinHash sketch. This buffered version better manages critical update operations thus significantly reducing the number of times the sketch needs to be rebuilt from scratch using expensive recovery queries. We prove that the buffered k-MinHash uses $O(k \log |U|)$ memory words per subset and that its amortized update time per insertion/deletion is $O(k \log |U|)$ with high probability. Moreover, our data structure can return the k-MinHash signature of any subset in $O(k)$ time, and this signature is exactly the same signature that would be computed from scratch (and thus the quality of the signature is the same as the one guaranteed by the static k-MinHash).
Analytical and experimental comparisons with the other, state- of-the-art global solutions for this problem given in [Bury et al.,WSDM'18] show that the buffered k-MinHash turns out to be competitive in a wide and relevant range of the online input parameters.

The poster is based on a joint work with Andrea Clementi and Luciano Gualà (University of Rome "Tor Vergata").

Modelling advection on distance-weighted directed networks

Francesco Zigliotto, Scuola Normale Superiore, Pisa

In this poster, we introduce a mathematical model to capture and describe the dynamics of the advection differential equation on distance-weighted directed graphs, with applications to various networked systems. The primary objective of our model is to generalize advection processes—which traditionally describe phenomena like fluid flow or traffic movement—by formulating them within the framework of discrete network structures.
Our approach begins with defining a set of essential properties, or axioms, that any discrete advection operator must satisfy. These axioms ensure that the operator reflects the fundamental characteristics of advection processes in continuous spaces, such as directional flow, conservation properties, and respect for edge weights based on distance. We demonstrate that, under these conditions, there exists an essentially unique operator that fulfills all prescribed axioms, providing a robust and mathematically sound foundation for discrete advection on directed graphs.
To validate our theoretical results, we present both analytical and numerical examples that showcase the operator's behavior under various configurations. As a practical demonstration, we apply the model to simulate traffic flow in a transportation network.

The poster is based on a joint work with Michele Benzi (Scuola Normale Superiore) and Fabio Durastante (University of Pisa).