

Analysing discursive communities and semantic networks on Twitter: an entropy-based approach

Tommaso Radicioni^{*1,2} Tiziano Squartini² Fabio Saracco²

* Corresponding author: tommaso.radicioni@sns.it

¹ Scuola Normale Superiore, P.zza dei Cavalieri 7, Pisa (Italy) ² IMT School for Advanced Studies, P.zza S. Ponziano 6, Lucca (Italy)

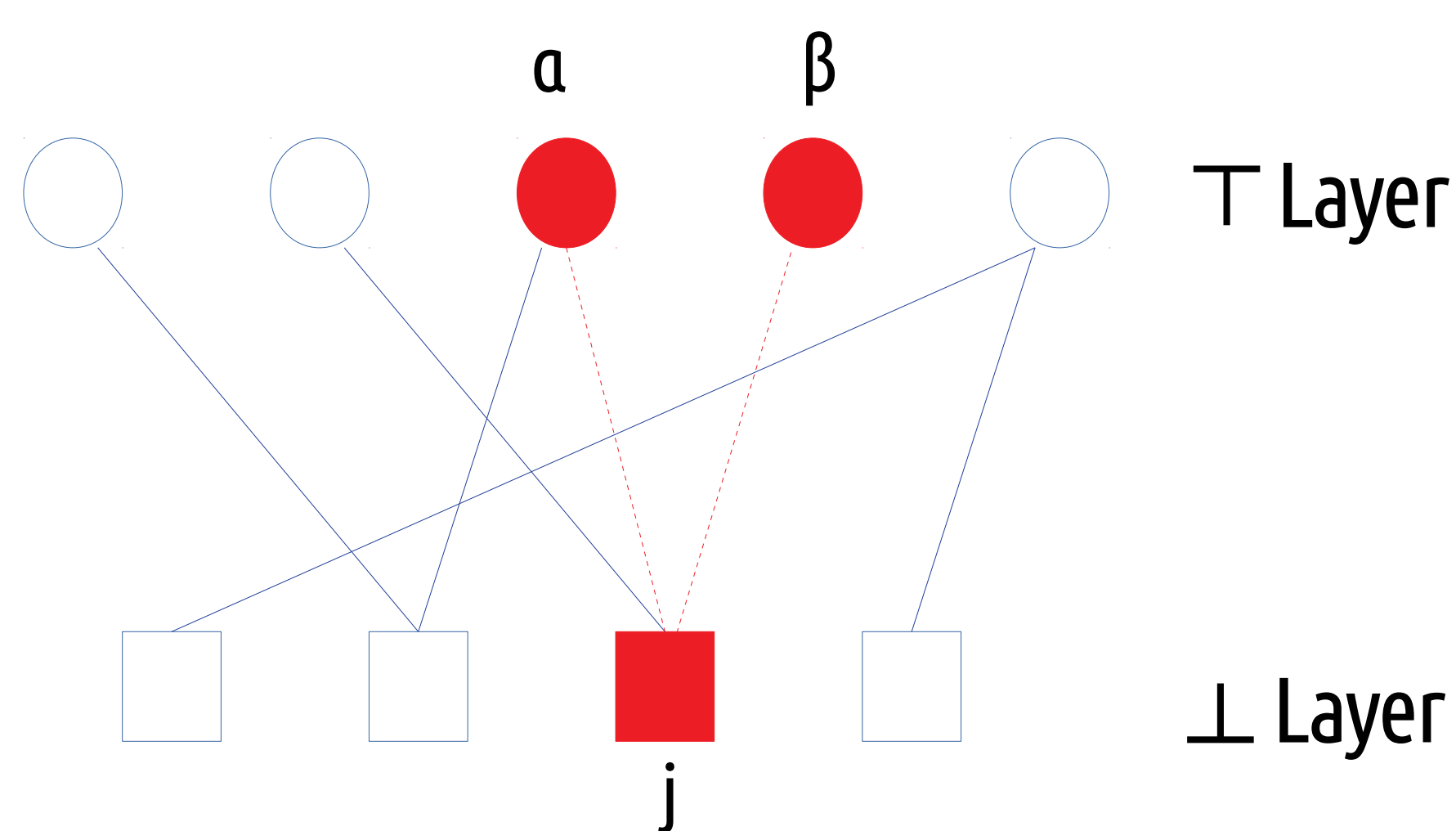
Introduction

Massive data availability in social media platforms allows to aggregate an enormous amount of information. Social media interactions needs to be filtered out to select only relevant information out of this data deluge. We propose an **entropy-based framework** to extract relevant semantic and social information from the online dynamics of Twitter discussion.

Methods

Communication process within Twitter stylized in two **bipartite structures**, i.e. (i) network of verified users retweeted by non-verified users and (ii) users connected with their published or shared hashtags. **Monopartite projections** are obtained with a three-steps procedure:

1. Measuring the **degree of similarity** of nodes α and β by counting the number of common neighbors $V_{\alpha\beta}$:

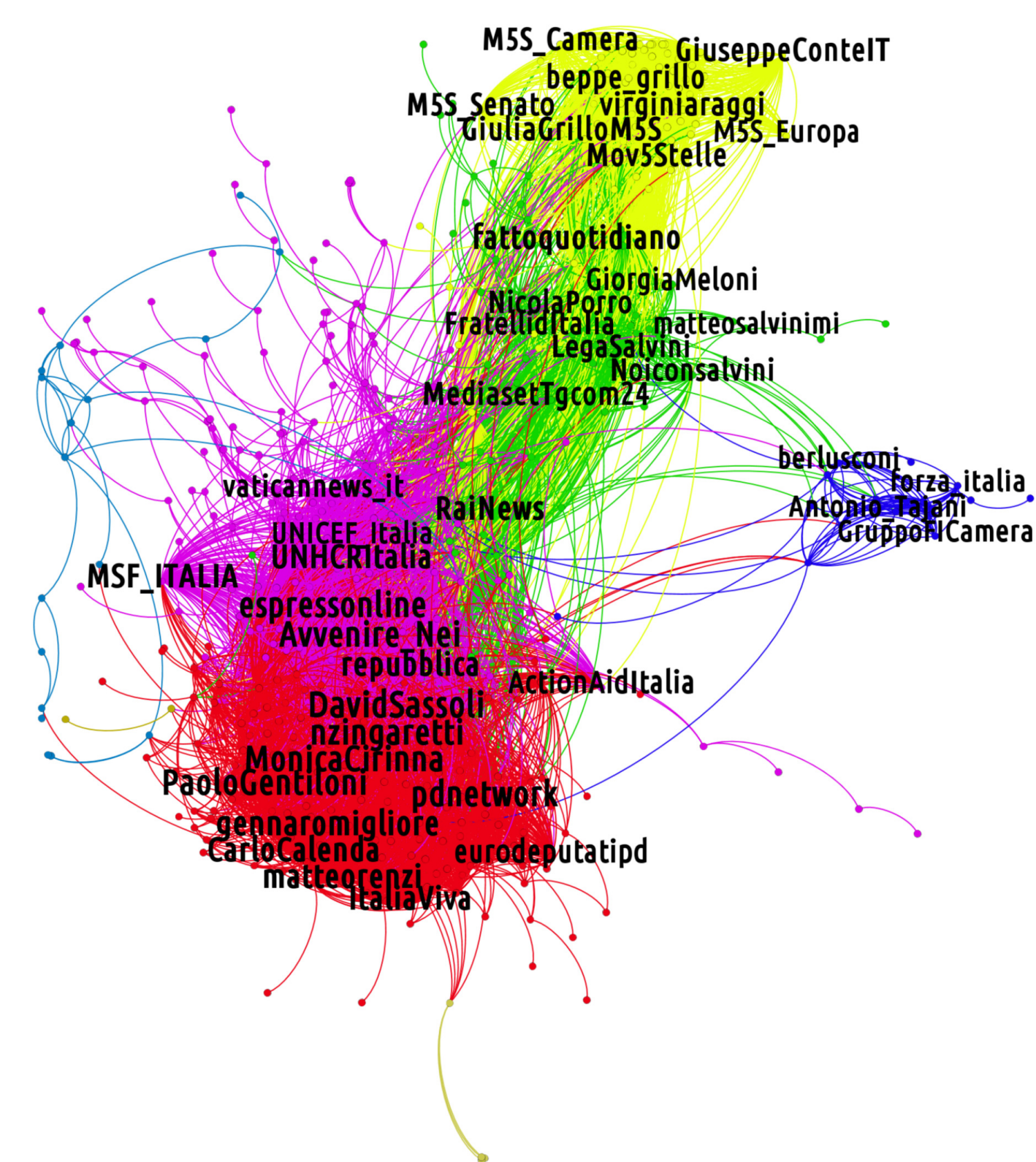


2. Quantifying the **statistical significance** of similarity between the two nodes through a benchmark **null model**, i.e. an entropy maximization procedure returning a connection probability $p_{\alpha j}$ from an ensemble of networks constrained on *average* (e.g. degrees of nodes belonging to both layers in **BiCM**);
3. Computing **p-values** of a **Poisson-Binomial distribution** as the probabilities of observing a number of V -motifs larger than the empirical one $V_{\alpha\beta}^*$, i.e. $\sum_{V_{\alpha\beta} \geq V_{\alpha\beta}^*} f(V_{\alpha\beta})$, and validating them through a procedure called **False Discovery Rate (FDR)** for testing simultaneously multiple hypotheses below a certain threshold (e.g. 0.01).

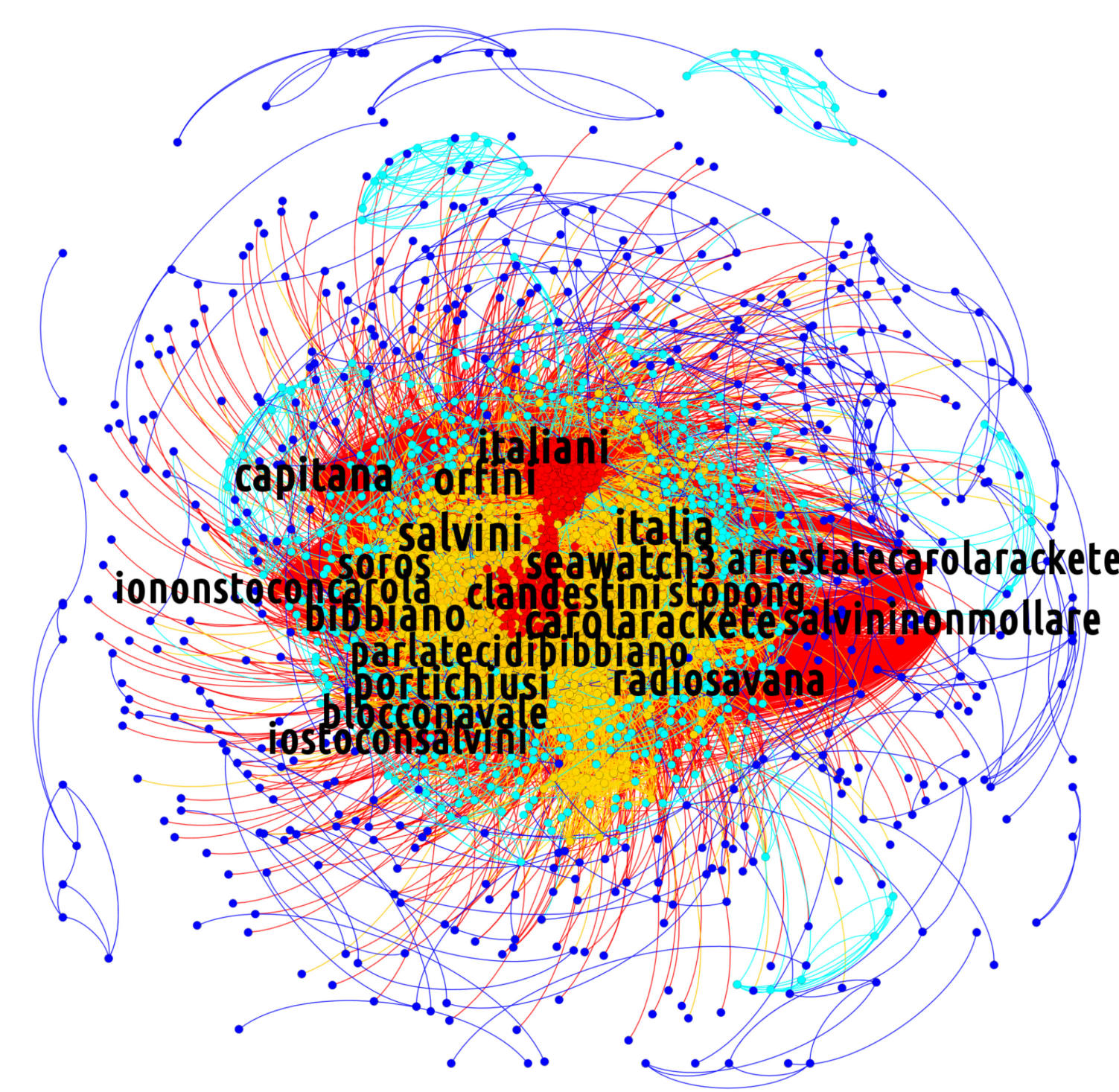
Results

Null models validation procedure act as a filter showing the non-trivial links in the original bipartite structure:

- Formation of **discursive communities** - a community detection algorithm is implemented to detect a **behaviour-driven** (rather than ideologically determined) picture of users **political leaning**;



- Construction of community-induced **semantic networks** - a conversational dynamics analysis spots out the presence of a **core-periphery** structure of hashtags showing a distinctive semantic approach in each community.



Key messages

Twitter interactions are studied throughout the lens of the analysis of the collective online dynamics during political discussions.

Constrained **Shannon entropy maximization** guarantees that the similarity between any two nodes is evaluated unbiasedly **without resorting upon any manual labelling** of either the users features or their sharing patterns.

Discursive communities are defined as *groups of Twitter users who share significantly similar retweeting patterns*. The most prominent contents circulating within them are analyzed trough **semantic networks** induced via hashtagging practices.

