How News Connects: Mapping the Semantic Topology of a Digital Newspaper

Sofia Sciangula, Andrea Agosta, Federico Carucci, Francesco Bertolotti

School of Industrial Engineering, LIUC - Carlo Cattaneo University, Castellanza, Italy
Intelligence, Complexity and Technology Lab (ICT Lab), LIUC - Carlo Cattaneo University, Castellanza, Italy
fbertolotti@liuc.it

Digital newsrooms continuously generate large volumes of information, which is not often analyzed as a networked system. This study models an Italian online newspaper, "Il Post", as a semantic ecosystem by representing its content as a bipartite article–tag network and projecting it onto a tag–tag network, to explore how pieces of information connect.

The resulting structure is highly asymmetric, with only a small group of tags – frequently belonging to specialized areas such as politics or economics – that functions as broad connectors across diverse topics. Despite overall sparsity, a giant connected component integrates nearly all areas, indicating both a coherent tag assignment from the sources and a navigable thematic landscape. Community detection identifies distinct thematic modules, corresponding to the major areas of public discourse.

In addition, the tag-tag network exhibits small-world characteristics: the clustering coefficient is substantially higher than that of comparable random or degree-preserving models, while the average shortest-path length is low and scales logarithmically with the number of nodes.

Preliminary conclusions show how the complexity of the world is mirrored by the tag-tag news network and how it has emergent small-world features. Future developments could be made with this framework, such as comparing different newspapers or studying its evolution in time.

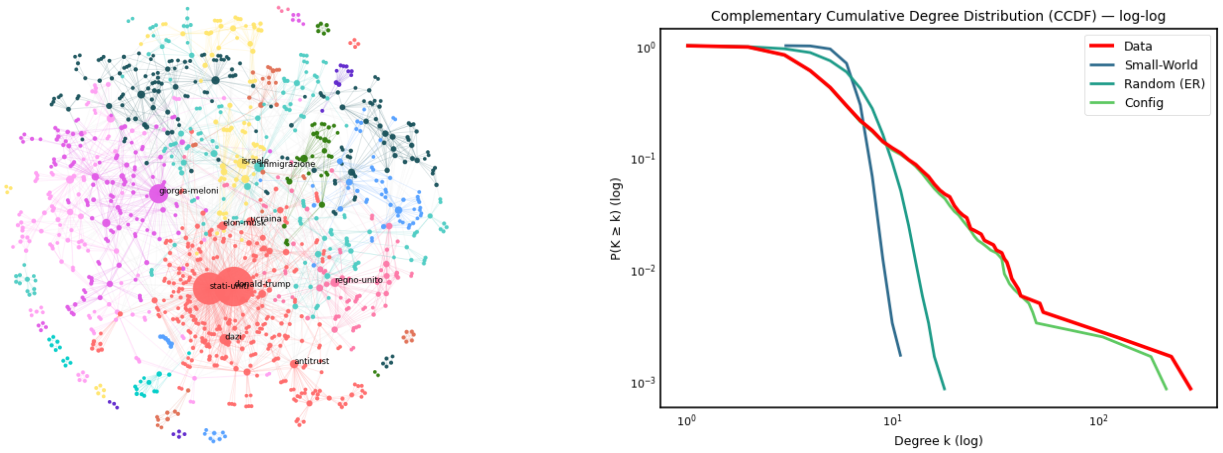

Figure 1: **Left:** Community structure of the weighted tag-tag network (using data from the year 2025 for the sake of a greater readability), illustrating the modular semantic landscape of the newspaper "Il Post". **Right:** Comparison of the complementary cumulative degree distributions for the real network and the reference models.