

# Distance Matters: Geo-social Metrics for Online Social Networks

Salvatore Scellato

## Abstract

In the recent years massive Online Social Networks (OSNs) such as Facebook, LinkedIn and Twitter have become increasingly popular, gathering millions of users and engaging them in the production, sharing and consumption of information over social links. These OSNs are increasingly becoming location-aware: they offer the opportunity to share geographic location in order to generate location-tagged information and to search for it. These new features open novel research directions which are largely unexplored, such as the design of new social applications and the improvement of existing large scale systems. Hence, it becomes important to investigate how geographic distance between individuals affects OSNs in order to deepen our understanding of these networks.

In this work we present a new approach for the analysis of networks with geographic information: we define new geo-social metrics which are able to quantify the geographic properties of the social ties across people. We describe a framework where network nodes are embedded in a metric space, in order to study the relationship between social connections and geographic distance. We define two new geo-social measures: a *node locality* metric, which quantifies how much a node is engaged with a local rather than global set of individuals, and a *geographic clustering coefficient*, which extends the standard notion of clustering by taking into account how much clusters of people are connected by short-range ties.

We apply our metrics to four different OSNs which provide location information for their users: we study two purely location-based social networking services (BrightKite and FourSquare), one blogging community (LiveJournal) and a social micro-blogging platform (Twitter). In particular, our results show how the new generation of location-based services is characterized by short-range friendship links among users, resulting in a vast proportion of them with high values of node locality. On the contrary, these patterns are not present in social networks which are less centered on user location: in LiveJournal users have connections with heterogeneous length and this effect is even greater in Twitter. Moreover, we find that location-based OSNs such as FourSquare and BrightKite tend to have more geographically confined triangles than social networks more focused on content production and sharing such as LiveJournal and Twitter, while standard clustering coefficient is not affected by this distinction.

These results indicate how taking into account geographic distance in these metrics provides insightful information for the design of systems and applications that could potentially exploit the underlying geographic and social structure of OSNs.