

Plenary Talk

About graph similarity algorithms

V. Blondel, L. Ninove, P. Van Dooren
CESAME, Université catholique de Louvain
Av. G. Lemaitre 4, B-1348, Louvain-la-Neuve, Belgium

Abstract

We introduce a concept of similarity between vertices of directed graphs. Let G_A and G_B be two directed graphs with respectively n_A and n_B vertices. We define a $n_A \times n_B$ *similarity matrix* \mathbf{S} whose real entry s_{ij} expresses how similar vertex i (in G_A) is to vertex j (in G_B): we say that s_{ij} is their *similarity score*. In the special case where $G_A = G_B = G$, the score s_{ij} is the similarity score between the vertices i and j of G and the square similarity matrix \mathbf{S} is the *self-similarity matrix* of the graph G . We show the use of this notion in the extraction of synonyms from dictionaries.

Other authors have defined a similar concept based on the fixed points of one of the following iterations

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}, \quad x_{k+1} = \frac{Ax_k + b}{\|Ax_k + b\|}, \quad x_{k+1} = \frac{Ax_k + b}{y^T(Ax_k + b)},$$

where the initial vector x_0 is given. We discuss the convergence properties of these different iterations and compare their properties.

References

- [1] V. Blondel, A. Gajardo, M. Heymans, P. Senellart, P. Van Dooren, *A measure of similarity between graph vertices: Applications to synonym extraction and web searching*, SIAM Review, to appear.
- [2] G. Jeh, J. Widom, *SimRank: A measure of structural-context similarity*, Proc. of the KDD2002 Conf., Edmonton (2002)
- [3] S. Melnik, H. Garcia-Molina and A. Rahm, *Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching*, Proc. 18th ICDE Conf., San Jose (2002).