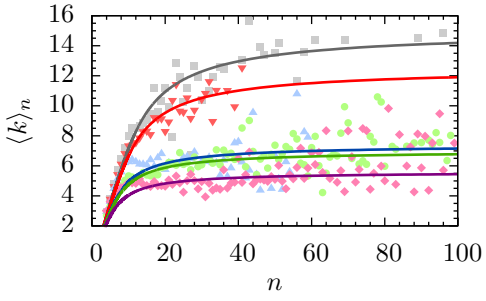
Structural preferential attachment of community structure and its relation to Dunbar's number

Jean-Gabriel Young, Laurent Hébert-Dufresne, Antoine Allard and Louis J. Dubé.
Département de physique, de génie physique et d'optique, Université Laval, Québec (Qc), Canada G1V 0A6

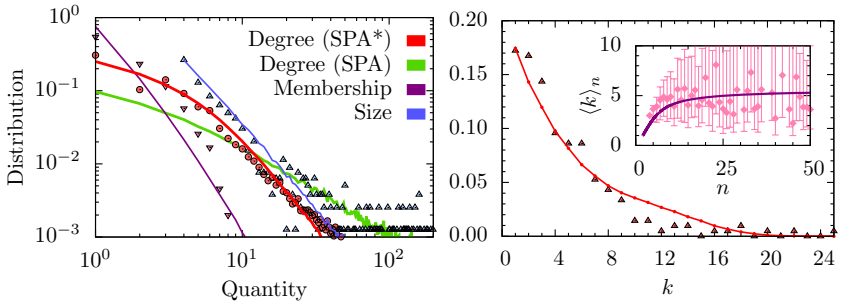
The social activity of individuals within communities are limited by their ability to maintain stable relationships with their peers. From a network perspective, this observation translates into empirical limits (Dunbar's number) on the maximal degrees that nodes can have within each of the communities to which they belong. It has been proposed that this constraint arises as a consequence of an individual's limited cognition resources. We show that such group behavior can also be understood as an emerging property of a simple system of two social mechanisms, independent of the actual nature of the network's nodes. Our idea is based on the simple assumption that each individual can, for every social group to which it belongs, develop connections and introduce new members. The resulting model accurately reproduces the limited internal degrees that are observed in real social networks. In fact, using our growth mechanism within a recently introduced structural preferential attachment (SPA) model [1], we reproduce with unprecedented accuracy the community structure, the degree distribution *and* the realistic internal structure of the communities of actual complex networks. This combined stochastic growth model yields an important additional insight into the community structure of networks: it suggests that vast, sparse, and therefore undetectable, communities are naturally occurring in social networks.

Submitted for an **oral contribution** to the International Workshop and Conference on Network Science (NetSci) held in Berkeley from June 2nd to June 6th 2014.

[1] Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A., and Dubé, L.J., Structural Preferential Attachment: Network Organization beyond the Link. *Phys. Rev. Lett.*, **107**:158702, 2011.



(a) **Dunbar's number:** Average number of connections $\langle k \rangle_n$ for a given individual within a group of size n in the arXiv cond-mat network (circa 2005), as detected by 5 popular algorithms. The prediction of our single parameter model are shown using solid lines. Plateau behavior is independent of the choice of algorithm, and thus of the *exact* definitions of communities.



(b) **Modelization of the Enron email network:** (left) Observed distributions (circle and triangles) of community structure related quantities for the Enron network, as detected by the greedy clique expansion algorithm. Simulated distributions (solid lines) are shown both for the SPA model and the combined SPA+local model (SPA*). The complete model only relies on 3 parameters tuned *independently*, using the scaling exponent of both the size and membership distributions, as well as the height of the plateaus shown in Fig. (a). (right) Internal degree distribution for large communities detected by the link clustering algorithm (triangles) and as predicted by our local growth model (line). The inset shows the raw dataset used to tune the growth model.