A new graph density
Version 1

Sergey Kirgizov [1]

1. Laboratoire Electronique, Informatique et Image (Le2i). CNRS : UMR6306 - Université de Bourgogne - Arts et Métiers ParisTech

Abstract For a given graph G we propose the non-classical definition of its true density: $\rho(G) = \mathcal{M}ass(G)/\mathcal{V}ol(G)$, where the $\mathcal{M}ass$ of the graph G is a total mass of its links and nodes, and $\mathcal{V}ol(G)$ is a size-like graph characteristic, defined as a function from all graphs to $\mathbb{R} \cup \infty$. We show how the graph density ρ can be applied to evaluate communities, i.e "dense" clusters of nodes.

0   Background and motivation

Take a simple graph $G = (V, E)$ with n nodes and m links. The standard definition of graph density, i.e. the ratio between the number of its links and the number of all possible links between n nodes, is not very suitable when we are talking about the *true density* in the physical sense. More precisely, by "the true density" we mean: $\rho(G) = \mathcal{M}ass(G)/\mathcal{V}ol(G)$, where the $\mathcal{M}ass$ of the graph G equals to total mass of its links and nodes, and the $\mathcal{V}ol$ is a size-like characteristic of G.

Consider again the usual graph density: $D = \frac{2m}{n(n-1)}$. Rewriting D in the "mass divided by volume" form, one obtain the following definitions of graph mass and volume:

$$\mathcal{M}ass_D(G) = 2m\,,$$
$$\mathcal{V}ol_D(G) = n\,(n-1)\,;$$

Note, that $\mathcal{V}ol_D(G)$ depends only of the number of nodes, so it is very rough estimation of the actual graph volume. Moreover, any function of the number of nodes (and the number links) will give somewhat strange results, because we neglect the actual graph structure in this way.

In the next section of this article we give a formal definition of the actual graph volume. For the moment, just take a look at the Fig. 1, where different graphs with 6 nodes and 6 links are shown. Intuitively, graph C is larger (more voluminous) than B and A. But it is not clear which graph is larger: A or B.
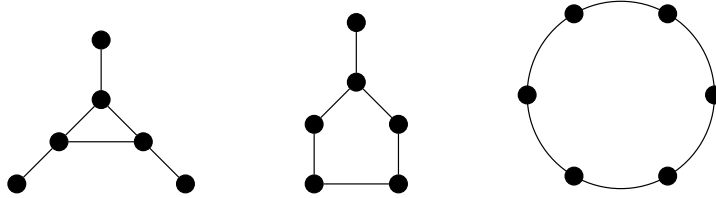

Figure 1: Examples of different graphs with 6 nodes and 6 links. These graphs have the same diameter, but their volumes are necessary the same.

1   True graph density

1.1 $\mathcal{M}ass(G)$

It seems a god idea to define $\mathcal{M}ass(G)$ as the total mass of its nodes and links. The simple way consists in assuming that the mass of one link (or node) equals to 1.

$$\mathcal{M}ass(G) = n + m \tag{MASS}$$

1.2 $\mathcal{V}ol(G)$

We cannot use any classical measure (e.g. Lebesgue-like) to define a volume of a graph G, because all measures are additive. Let us explain why the additivity is bad. Observing that G is the union of its links and nodes, and assuming that the volume of a link (node) equals to one, we obtain:

$$\mathcal{C}lassical\mathcal{V}ol(G) = n + m\,,$$

where m is the number of links in G, and n equals to the number of nodes. The graph structure disappears again, and we should find "another definition of volume".

A clever person can develop a notion of "volume" for any given metric space. Since any graph can be regarded as a metric space, we can use this as a solution of our problem.

Here we briefly describe how Feige in his paper [2] defined the volume of a finite metric space (S, d) of n points. A a function $\phi : S \to \mathbb{R}^{n-1}$ is a *contraction* if for every $u, v \in S$, $d_{\mathbb{R}}\big(\phi(u) - \phi(v)\big) \le d(u, v)$, where $d_{\mathbb{R}}$ denotes usual Euclidean distance between points in $\mathbb{R}^{n-1}$. The Fiege's volume $Vol\big((S, d)\big)$ is the maximum $(n-1)$ dimensional Euclidean volume of a simplex that has the points of $\{\phi(s) | s \in S\}$ as vertices, where the maximum is taking over all contractions $\phi : S \to \mathbb{R}^{n-1}$. Sometimes in order to calculate Fiege's volume, we need to modify the original metric. Abraham *et al.* deeply studied Fiege-like embeddings in [1].

Another approach is to find a *good* mapping $g : S \to \mathbb{R}^{n-1}$, trying to preserve original distances as much as possible, and calculate the $Vol\big((S, d)\big)$ as the volume of convex envelop that contains all $\{g(s) | s \in S\}$. The interested reader can refer to the Matoušek's book [3], which gives a good introduction into such embeddings. But we should note that not all finite metric spaces can be embedded into Euclidean space with exact preservation of distances.

In this paper we chose another approach: instead of doing approximative embeddings, we compute the "volume" directly. First of all, let us introduce some *natural* properties that must be satisfied by the graph volume. A graph volume is a function from set of all graphs $\mathcal{G}$ to $\mathbb{R} \cup \infty$:

$$\mathcal{V}ol : \mathcal{G} \to \mathbb{R} \cup \infty\,,$$

Note that our volume has no such parameter as dimension. The absence of dimension allows us directly compare volumes of any two graphs.

Let the volume of any complete graph be equal to 1:

$$\mathcal{V}ol(K_x) = 1 \tag{I}$$

Then, for any disconnected graph, denoted by $G_{\bullet\bullet}$, let the volume be equal to infinity:

$$\mathcal{V}ol(G_{\bullet\bullet}) = \infty \tag{II}$$

Intuitively, here one can make an analogy with a gas. Since gas molecules are "not connected", they fill an arbitrarily large container in which they are placed.

When we add a new edge between two existed vertices, the new volume (after edge addition) cannot be greater than the original volume:

$$\mathcal{V}ol(G) \geq \mathcal{V}ol(G + e) \tag{III}$$

When we add a new vertex v^1 with degree 1, the new volume cannot be less than the original one:

$$\mathcal{V}ol(G) \leq \mathcal{V}ol(G + v^1) \tag{IV}$$

For a given graph $G = (V, E)$ the eccentricity $\epsilon(v)$ of a node v equals to the greatest distance between v and any other node from G:

$$\epsilon(v) = \max_{u \in V} d(v, u) \,,$$

where $d(v, u)$ denotes the length of a shortest path between v and u.

Finally, we define the volume of a graph G as a product of all eccentricities:

$$\mathcal{V}ol(G) = \sqrt[|V|]{\prod_{v \in V} \epsilon(v)} \tag{VOLUME}$$

Obviously properties I, II and III hold for this definition. But IV is needed to be proved or disproved. Reconsidering graphs from Fig. 1, we have $\mathcal{V}ol(A) = \sqrt[6]{3^3 2^3} \approx 2.45$, $\mathcal{V}ol(B) = \sqrt[6]{3^3 2^3} \approx 2.45$ and $\mathcal{V}ol(C) = \sqrt[6]{3^6} = 3$.

2 Possible applications

Quality of communities   Consider two graphs A and B. We say that A is *better* than B if and only if $\rho(A) > \rho(B)$. Using this notion one can define a quality of graph partition.

The volume of finite metrics spaces   Our approach can be applied to calculate the "volume" of any finite metric space (S, d):

$$\mathcal{V}ol\big((S, d)\big) = \sqrt[|S|]{\prod_{s \in S} \epsilon(s)} \,,$$

where $\epsilon(s) = \max_{p \in S} d(s, p)$.

References

[1] I. ABRAHAM, Y. BARTAL, O. NEIMAN, AND L. J. SCHULMAN, *Volume in general metric spaces*, in Proceedings of the 18th annual European conference on Algorithms: Part II, ESA'10, Berlin, Heidelberg, 2010, Springer-Verlag, pp. 87–99.

[2] U. FEIGE, *Approximating the bandwidth via volume respecting embeddings*, Journal of Computer and System Sciences, 60 (2000), pp. 510 – 539.

[3] J. Matoušek, *Lectures on Discrete Geometry*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.