

Central loops in random planar graphs

Benjamin Lion, Marc Barthelemy

▶ **To cite this version:**

Benjamin Lion, Marc Barthelemy. Central loops in random planar graphs. 2017. cea-01502150

HAL Id: cea-01502150
https://cea.hal.science/cea-01502150

Preprint submitted on 5 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1611.03232v3 [physics.soc-ph] 27 Mar 2017

Central loops in random planar graphs

Benjamin Lion and Marc Barthelemy*
Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France

Random planar graphs appear in a variety of context and it is important for many different applications to be able to characterize their structure. Local quantities fail to give interesting information and it seems that path-related measures are able to convey relevant information about the organization of these structures. In particular, nodes with a large betweenness centrality (BC) display non-trivial patterns, such as central loops. We first discuss empirical results for different random planar graphs and we then propose a toy model which allows us to discuss the condition for the emergence of non-trivial patterns such as central loops. This toy model is made of a star network with N_b branches of size n and links of weight 1, superimposed to a loop at distance ℓ from the center and with links of weight w. We estimate for this model the BC at the center and on the loop and we show that the loop can be more central than the origin if $w < w_c$ where the threshold of this transition scales as $w_c \sim n/N_b$. In this regime, there is an optimal position of the loop that scales as $\ell_{opt} \sim N_b w/4$. This simple model sheds some light on the organization of these random structures and allows us to discuss the effect of randomness on the centrality of loops. In particular, it suggests that the number and the spatial extension of radial branches are the crucial ingredients that control the existence of central loops.

PACS numbers: 89.75.Fb, 89.75.-k, 05.10.Gg and 89.65.Hc

INTRODUCTION

Random planar graphs – random graphs that can be drawn on the 2d plane with no edge crossing [1] – pervade many different fields from abstract mathematics [2, 3], to theoretical physics [4], botanics [5, 6], geography and urban studies [9]. In particular, planar graphs are central in biology where they can be used to describe veination patterns of leaves or insect wings and which display an interesting architecture with many loops at different scales [5–8]. In the study of urban systems, planar networks are extensively used to represent, to a good approximation, various infrastructure networks [9] such as transportation networks [10] and streets patterns [11–29]. Understanding the structure and the evolution of these networks is therefore interesting from a purely graph theoretical point of view, but could also have an impact in different fields where these structures are central.

Most previous studies characterize different aspects of these graphs, either purely topological (degree distribution, clustering, etc.) or geometrical (angles, segment length, face area distribution, etc.). Due to spatial constraints, most local information such as the degree distribution, the clustering or the assortativity have however a trivial behavior [9]. In addition the important information about these random planar graphs is in fact not in their adjacency matrix only but also in their geometry described by the spatial distribution of nodes and relevant meaures should combine topology and geometry. Despite the large number of studies on these graphs, there is still a lack of a non-local high-level metrics that allow for understanding and comparing these graphs with each other. However, a promising direction is given by path-based quantities such as the simplicity [27, 30] or as we will discuss here, the betweenness centrality (BC) [31]. The BC was introduced to quantify the importance of a node (or an edge) in a network, but it also proved to be a very interesting tool in the study of random planar graphs. Already in [17], interesting spatial patterns of nodes with large BC were observed. More recently, it has been shown that the most salient aspects of the structural changes during the evolution of the street network of Paris [26] is revealed by the spatial distribution of the nodes with the largest BC. These different results therefore point to the fact that high centrality nodes form non-trivial patterns, and among them, the emergence of loops. All these different results point to the fact the BC, which is relatively simple, is a good candidate for monitoring and understanding the organization of random planar graphs. In this study we will focus on the apearance of loops made of links with large BC and we will propose a simple toy model that allows us to discuss the conditions for the appearance of such patterns.

The emergence of rings in largely urbanized areas is a common fact and the study presented here gives a topological light on this phenomenon. Our study echoes previous work where congestion effect at a central hub could be so high that avoiding the ring is beneficial [32, 33]. Here, in contrast, we do not take into account congestion and discuss the conditions necessary for a loop to become more interesting in terms of time cost.

THE BC FOR PLANAR GRAPHS

Basic results on planar networks can be found in any graph theory textbook (see for example [1] and for useful algorithms see [34]) and we will very briefly recall the definition of these objects. Basically, a planar graph is a

graph that can be drawn in the plane in such a way that its edges do not intersect. Not all drawings of planar graphs are without intersection and a drawing without intersection is sometimes called a plane graph or a planar embedding of the graph. In real-world cases, these considerations actually do not apply since the nodes and the edges represent in general physical objects. More precisely, we will focus here on the case of planar graphs embedded in 2d space, which typically describes systems such as the road and street network.

Definition of the BC and variants

The betweenness centrality counts the fraction of shortest paths going through a given node (or link) and is given by [31]

$$g(v) = \mathcal{N} \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

where v is a node, σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ those paths going through v. In general the summation is on $s \neq t$ and $s \neq v$, $t \neq v$ and this is the convention that we will adopt in this paper. The constant $\mathcal{N}$ is the normalisation and we will use here $\mathcal{N} = 1/(N-1)(N-2)$ which counts the number of pairs and ensures that $g(v) \in [0, 1]$. For edges, the definition of the BC is similar to Eq. 1.

It is important to stress here that the BC could in fact be defined for any type of paths. The most common choice is the shortest path but we will use the more general case of weighted shortest path, which corresponds to the quickest path if the weight of a link represents time. For numerical calculations, we implemented the now standard algorithm of Brandes [35].

Regular lattice

In a one-dimensional lattice of size n, the BC of a site $0 \leq x \leq n$ is given by

$$g(x) = x(n - x) \tag{2}$$

and the maximum thus corresponds to the barycenter of all nodes (see Fig.1) (in the two-dimensional case we obtain a similar behavior). When we introduce disorder - by removing or rewiring links - the BC becomes important at nodes that can be far away from the barycenter (see Fig.1). In the extreme case where space doesn't play a role anymore such as in scale-free networks, the average BC per degree classes $g(k)$ scales as [36]

$$g(k) \sim k^{\eta} \tag{3}$$

where η is an exponent that depends on the structure of the graph. Even if there are fluctuations around this scaling it shows here that essentially the degree controls the BC in these graphs.

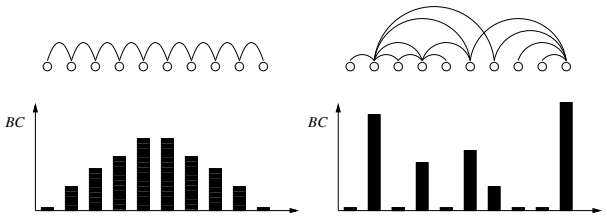


FIG. 1: Betweenness centrality for a one-dimensional lattice. (Left) When there is no disorder, the barycenter is the most central nodes. (Right) In the case of a disordered network, degree becomes relevant and the most central nodes have large degrees.

Percolation: giant component

A simple way to construct a random planar graph is to consider a regular lattice where each link has a probabilty f to be removed (and $p = 1 - f$ to be present). Above the percolation threshold ($p \geq p_c$), the system displays a giant component which connects a non-zero fraction of the nodes. We can study the BC on this giant component and filter them for different threshold g^*: we keep only links with centrality g such that $g > g^*$. We show in Fig. 2 the set of links that belong to the giant component and with BC larger than g^* and represent the BC with a color code (from dark blue to yellow). and we observe that the set of most central links forms a non-trivial

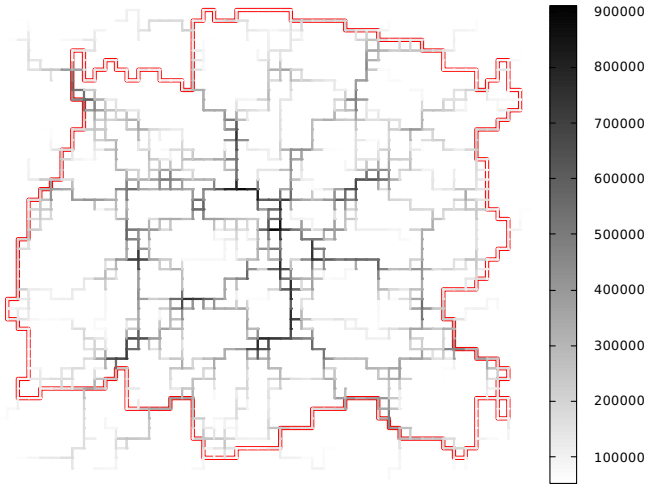


FIG. 2: Links belonging to the giant component obtained for percolation with $f = 0.15$ and with normalized BC larger than $g^* = 0.05 g_{max}$. We observe that the links with a very large BC form a non-trivial pattern and are not necessarily close to the center. The largest loop is here highlighted with a double line.

pattern where the distance to the center is not the main determinant (Fig. 2). In particular, we observe the presence of very central links that are not close to the center and that depend on the particular disorder configuration.

We can go further in the analysis of the structure of the percolating cluster by analyzing the ratio

$$\eta = \frac{g(r,\theta)}{\max_{r'<r,\theta'\in[0,2\pi]} g(r,\theta')} \tag{4}$$

which compares the BC at one point with the maximum BC of nodes in the region closer to the origin. For a percolating cluster obtained at $p = 0.8$ (well above the percolation threshold) and on a lattice 100×100, we observe a very broad distribution of η. For values larger than one we obtain on average $\overline{\eta} \simeq 3$ and a very large dispersion of order 10^3. We can observe the points for which we have a ratio $\eta > 1$ and plot (Fig. 3) the distribution of the distance to the center for these points (normalized by the maximum distance d_{max}). This figure 3 shows that the location of nodes with a very large BC (at least larger than the BC of the nodes closer to the center) can be of order the system size. This shows that – depending on the disorder – the 'central' area composed of the geometrical center and its surroundings are composed of nodes with a relatively small BC. This reinforces the need to understand in which cases the monotoneous decrease of the BC with the distance to center can be strongly modified by fluctuations.

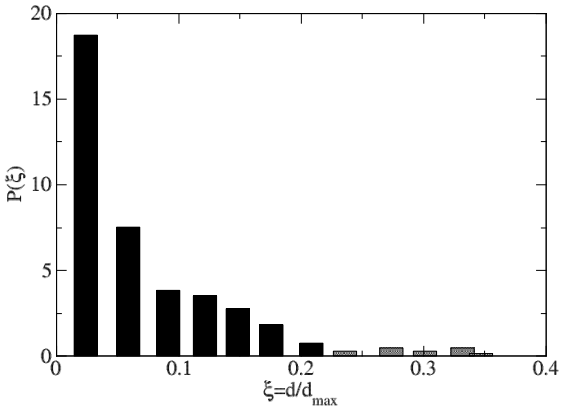


FIG. 3: Distribution of the distance (normalized by the maximum distance on the lattice) to the center for nodes with a ratio $\eta > 1$. These results were obtained for $p = 0.8$ on a 100×100 lattice and averaged over 30 configurations.

Real-world planar graphs

Streets and roads form a network where nodes are intersections and links are segment roads, and which is planar (or almost planar, to a good approximation). This network is now fairly well characterized and due to spatial constraints, the degree distribution is peaked, the clustering coefficient and assortativity are large, and most of the interesting information lies in the spatial distribution of betweeenness centrality [9]. Many studies [13–26, 29] considered different aspects of this network and observed non-trivial structures in the BC spatial distribution. In particular, in [17] it has been observed that the distribution of the BC can display non-trivial spatial patterns and in [26] the authors showed that during the evolution of the street network of Paris (France) most 'standard' measures were unable to detect the important structural changes that occurred in the 19^{th} century, while in contrast, the spatial distribution of the BC displayed dramatic changes.

Using the road network obtained from city extracts (the data has been obtained from the Mapzen website [37]), we compute the BC distribution for different cities shown in Fig. 4. For all these real-world cases, we observe

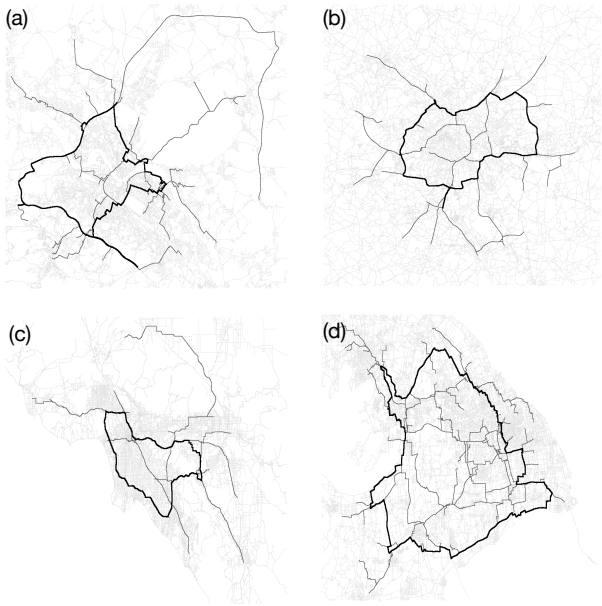


FIG. 4: We show for real-world networks the links e with a BC larger than a certain threshold $g(e) > g^*$ and we highlight the largest loop. (a) Dresden, Germany ($g^* = 0.11$). (b) Paris, France ($g^* = 0.315$). (c) Los Angeles, USA ($g^* = 0.05$). (d) Shanghai ($g^* = 0.07$).

that indeed non-trivial structures appear and in particular we observe the appearance of loops made of central links and of different sizes. We can test the stability of these loops, by filtering these networks for different values of the BC threshold g^* and compute the perimeter of the main loop. The results for Dresden, Los Angeles, and Paris are shown in Fig. 5. We observe on this plot the presence of various plateaus at intermediate values of g^* suggesting that these loops are indeed very central

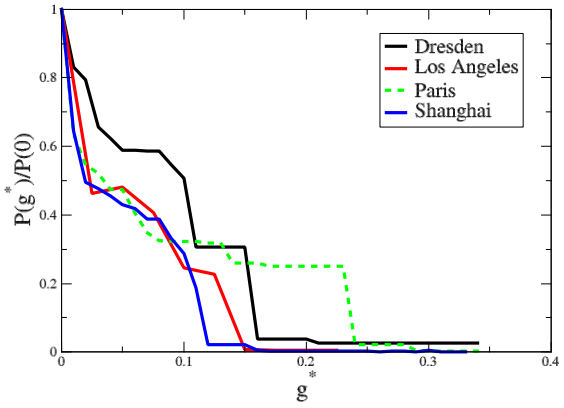


FIG. 5: Perimeter $P(g^*)$ of the main loop (normalized by perimeter of the loop at $g^* = 0$) for the road network of Dresden, Los Angeles, and Paris.

and stable.

In general, boundary effects can be important and can affect the measures done on spatial networks [38]. In general, the choice of boundaries has an impact on quantities such as the BC [39] and we briefly discuss this problem here. We measure the area enclosed in the largest loop on the same network but at different scales (ie. with different boundaries, going from central Paris to almost the whole Ile-de-France region) and the results are presented in Fig. 6. In this figure, we observe that at least the area

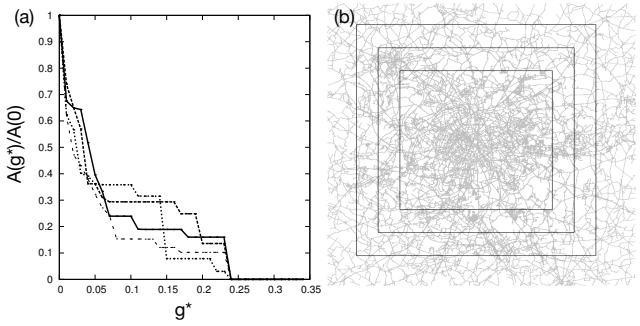


FIG. 6: (a) Normalized area $A(g^*)$ defined by the largest loop for different boundary conditions on the Paris road network. The lowest curve corresponds to the largest size and for decreasing size the curves are shifted to larger values of the area. (b) Different boundaries corresponding to the curves of (a).

of the largest loop remains relatively stable. Further systematic studies are however certainly needed in order to understand which patterns are stable and which ones are not, and what are the conditions on the boundaries in order to ensure stability of the main spatial patterns.

Summary: stylized facts

These different examples discussed above show that the introduction of disorder in planar graphs induce in general the formation of non-trivial structures made of links with a large BC. In particular, we observe the appearance of loops made of links that can have a BC larger than the barycenter. In other words, disorder can invert the typical behavior observed for regular lattice where the BC is decreasing monotonously from the barycenter. In the following, we propose a toy model which allows to discuss and to understand under which conditions a loop can become more central than the spatial center.

THEORETICAL APPROACH: A TOY MODEL

As discussed above, we observe that non-trivial objects such as loops can be very central in random graphs. It is important to understand the formation of these structures and the conditions for their existence. In particular, it seems that randomness can induce very large perturbation in the spatial distribution of the BC and where the barycenter is not the most central node. Equivalently, the BC could not be a simple decreasing function of the distance to the barycenter anymore. In order to understand this phenomenon, we propose here a simple toy model. We first construct a star network composed of N_b branches, where each branch is composed of n nodes. We then add a loop at distance ℓ from the center (see Fig. 7 for a sketch of this graph). We also consider here a more general case where the links are weighted and in this simplified model we assume that links have a weight equal to one and the loop segment between two consecutive branches has a weight given by w. The purely topological case then corresponds to the case $w = 1$. We then compute the BC using weighted shortest paths. This generalization allows us to discuss for example the impact of different velocities on a street network. In this case, w can be seen as the time spent on the segment and the weighted shortest path is then the quickest path.

Here, we want to discuss under which conditions the loop will be more central than the 'origin' at the center in this simplified network. Intuitively, for very large w, it is always less costly to avoid the loop, while for $w \to 0$, loops are very advantageous. The two main quantities of interest are therefore the centrality at the center denoted by $g_0(\ell, n, w)$ and the centrality, denoted by $g_C(\ell, n, w)$, at the intersection C of the branch and the loop. We then compute the difference $\delta g = g_0 - g_C$ and will study under which condition it can be negative.

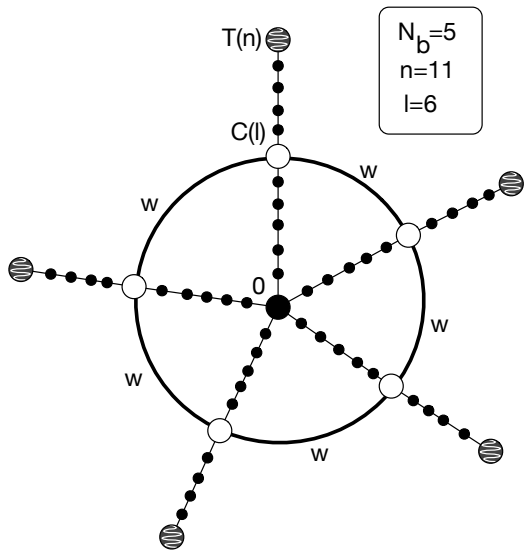


FIG. 7: Representation of the toy model discussed here. The number of branches is here $N_b = 5$, the number of nodes on each branch is $n = 11$ and the loop is located at a distance $\ell = 6$ from the center 0. The node C is at the intersection of a branch and the loop and T is the terminal node of a branch.

Exact and approximated formulas

The interest of this toy model lies in the fact that we can estimate analytically the BC for the center $g_0(\ell, w)$ and for the intersection nodes on the loop $g_C(\ell, w)$. Formally we can write these quantities as

$$\begin{aligned} g_0(\ell, n, w) &= g_0(\ell, n, \infty) - (a_1^0 + a_2^0 + a_3^0) \\ g_C(\ell, n, w) &= g_C(\ell, n, \infty) + (a_1^C + a_2^C + a_3^C) \end{aligned} \quad (5)$$

where the a_x^i are positive. We distinguish two parts in these centralities. First, we estimate the BC when there is no loop which is represented by the case where $w \to \infty$. This part is modified by the presence of the loop that under certain conditions can be more interesting for connecting pairs of nodes. We can understand the signs in Eq. 5, by noting that the presence of the loop will decrease the centrality at the center and increase the centrality at C. The different terms a_i^x (where $x = 0, C$ and $i = 1, 2, 3$) count the paths (that avoid 0) connecting two nodes that lie on different parts of their branch. We divide the nodes on a branch in two parts - the lower part comprises all nodes that are 'below' the loop $0 < s < \ell$ and the upper part is the rest $\ell < s \leq n$. When both nodes are on the upper part of the branches we obtain a_1^x; the paths connecting an upper part to a lower part are described by a_2^x and when both nodes lie on a lower part, we obtain the coefficient a_3^x. For more details and the calculation of these coefficients, we refer to the appendix.

The exact expressions for the centralities g_0 and g_C are however difficult to handle analytically, essentially because they are expressed as sums of complicated arguments (see appendix). In order to derive analytical predictions we will propose in the following a simple approximation scheme that allows to obtain the correct scalings for the most important quantities.

In the derivation of the exact expression of the centralities Eq. 5, we have to distinguish different cases according to the value of

$$\chi \equiv \min\left(\frac{N_b - 1}{2}, \left[\frac{2\ell}{w}\right]\right) \quad (6)$$

compared to $j-1$ (the brackets $[\cdot]$ denote here the integer part, ie. the lowest nearest integer) which denotes the number of loop segments between the first branch and the branch j. This essentially amounts to compare the cost of the path between a node on the lower part (with $0 < s < \ell$) of the first branch B_1 to a node on the lower part ($0 < t < \ell$) of another branch B_j. If $[2\ell/w] > j - 1$ the cost of the path which goes through 0 is larger than going directly via the loop (given by $(j-1)w$) and therefore produces a negative contribution to g_0. We see that this discussion allows to distinguish for a given value of w 'near' from 'far-away' branches (Fig. 8). The nearest branches are then defined by the condition $j - 1 \leq \chi$ and the remote branches by $\chi < j - 1 \leq (N_b - 1)/2$ (for simplicity we assume here that N_b is odd and by symmetry we can discuss only one half for the branches from $j = 2$ to $j = (N_b - 1)/2$). We will then use the

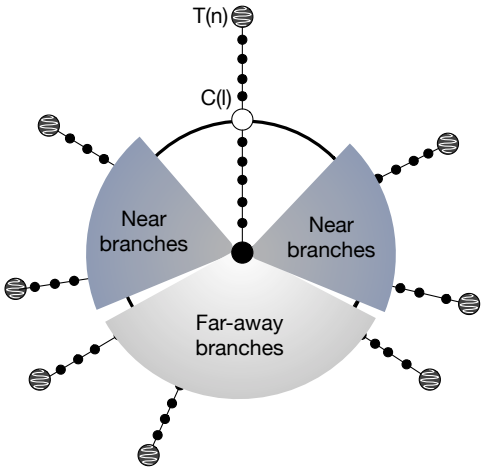


FIG. 8: Schematic representation of the approximation used to compute the centrality $g_0(w)$ at the center 0.

following simplification: we will assume that for the χ near branches, going through the center is always the best choice for $s < \ell$ and $t < \ell$ while for all other paths (from or to an upper path) it is beneficial to go through the loop. This leads then to

$$g_{near} = \chi \frac{(\ell - 1)(\ell - 2)}{2} \quad (7)$$

For the $(N_b-1)/2-\chi$ far-away branches, we consider that for all nodes $s,t \in [\![1,n]\!]$, the paths are going through the center leading to

$$g_{far} = \left(\frac{N_b-1}{2} - \chi\right) n^2 \tag{8}$$

Taking into account the factor 2 for not counting twice the same path, we obtain for $g_0(w) = N_b(g_{near} + g_{far})$ the following expression

$$g_0(w) \approx N_b \left\{ \left(\frac{N_b-1}{2} - \chi\right) n^2 + \chi \frac{(\ell-1)(\ell-2)}{2} \right\} \tag{9}$$

We note that this approximation recovers both exact limits

$$g_0 \simeq \begin{cases} N_b \frac{(N_b-1)}{2} \frac{(\ell-1)(\ell-2)}{2} & \text{for } w \to 0 \\ n^2 N_b \frac{(N_b-1)}{2} & \text{for } w \to \infty \end{cases} \tag{10}$$

In the following it will also be useful to consider the limit $\ell, n \to \infty$ with $x = \ell/n$ fixed which gives for $g_0(x,\chi) = g_0(\ell,n,w)/n^2$ (up to terms of order $1/n$)

$$g_0(x,\chi) \approx N_b \left\{ \left(\frac{N_b-1}{2} - \chi\right) + \frac{1}{2}\chi x^2 \right\} \tag{11}$$

where the only dependence on w is now encoded in χ, hence the change of argument for clarity.

We can produce the same type of arguments for the BC on the loop. First the value without the loop is easy to compute and we obtain

$$g_C(\ell,n,w=\infty) = (n-\ell)\left[\ell + n(N_b-1)\right] \tag{12}$$

which simply counts the number $(n-\ell)$ of nodes 'above' C and all the others (C being excluded). Similar arguments as above then give the following result (we also changed here the argument of the function from w to χ)

$$\begin{aligned} g_C(\ell,n,\chi) &= g_C(\ell,n,w=\infty) \\ &+ 2\chi\Big[(n-\ell+1)(\ell-1) + \frac{(\ell-1)(\ell-2)}{2}\Big] \\ &+ \frac{\chi(\chi-1)}{2}\Big[(n-\ell+1)^2 \\ &+ 2(n-\ell+1)(\ell-1) + \frac{(\ell-1)(\ell-2)}{2}\Big] \end{aligned} \tag{13}$$

where χ is given by Eq. 6. In particular the term proportional to χ counts all the paths between the lower part of the branch containing C and all the nodes of a branch close enough. The second term (proportional to $\chi(\chi-1)$) counts the paths going from a branch B_j with $j \in [1,\chi-1]$ to the other branches $j' = 1,2,\ldots,j-1$. The sum of all these contributions gives the factor $\chi(\chi-1)/2$. The counting factor is not trivial here and comes from evaluating all the paths from a node s in a branch j to a node t on a branch j' (j and j' are different from 1) such that

$$s + t > |\ell - s| + |\ell - t| + w\Delta j \tag{14}$$

The left hand side of this inequality corresponds to the distance through the center and $w\Delta j$ is the distance on the loop (for the exact expression of the centrality and how to recover this approximate formula, we refer the interested reader to the appendix).

Similarly to the case of the BC at 0, it will be convenient for analyzing these expressions to consider the limit $n, \ell \to \infty$ such that $\ell/n = x$. Up to terms of order $1/n$ we then obtain for $g_C(x,\chi) = g_C(\ell,n,w)/n^2$

$$\begin{aligned} g_C(x,\chi) &= (1-x)(x+N_b-1) \\ &+ 2\chi x(1-\frac{x}{2}) \\ &+ \frac{\chi(\chi-1)}{2}(1-\frac{x^2}{2}) \end{aligned} \tag{15}$$

We show in the figure 9 the comparison of the exact result with the approximations developed here. For large values of ℓ the approximation is not excellent and can certainly be improved. However as we will show in the following, our simple approximations allow to understand and to predict the correct scaling for the important quantities ℓ_{opt} and w_c.

Threshold value of w and optimal ℓ

The fundamental quantity that we wish to understand is the difference $\delta g(x,\chi) = g_0(x,\chi) - g_C(x,\chi)$ given by Eqs. (11,15). We first plot this quantity versus ℓ for different values of w and we observe the result shown in Fig. 10 This result shows that for w sufficiently small, δg can be negative. This demonstrates the existence of a threshold value w_c such that at $w = w_c$ the minimum is $\min_\ell \delta g = 0$. For $w < w_c$, the minimum of δg is negative and we can define an optimal value ℓ_{opt} which corresponds to this smallest value of δg. The quantity ℓ_{opt} thus gives the position of loop that maximizes the difference between the BC of the loop and the center.

In order to estimate this optimal value ℓ_{opt}, we note (using the expression Eq. 6 for χ) that the difference $\delta g(x,\chi)$ gives

$$\delta g(x,\chi) = \begin{cases} \delta g(x, \frac{2\ell}{w}) & \text{for } \ell \in [0, \frac{(N_b-1)w}{4}] \\ \delta g(x, \frac{N_b-1}{2}) & \text{for } \ell \in [\frac{(N_b-1)w}{4}, 2n] \end{cases} \tag{16}$$

In order to discuss to estimate analytically both the threshold w_c and the optimal value ℓ_{opt}, we will use equations Eqs. (11,15) and study the approximate difference

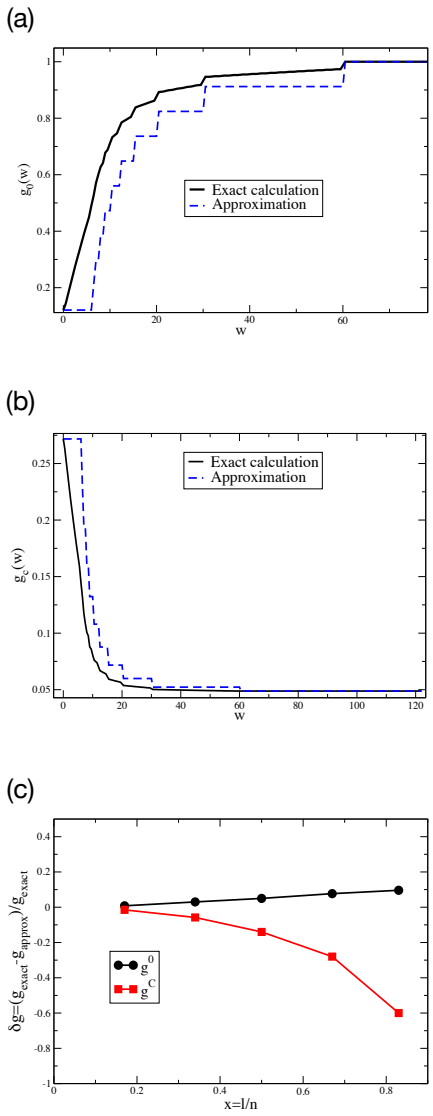


FIG. 9: Comparison between the exact result and the approximation for $g_0(w)$ (a) and $g_C(w)$ (b) (the BC are here normalized). The parameter values are here $N_b = 21$, $n = 60$ and $\ell = 30$. (c) Relative error between the exact value and approximation for g_0 and g_C for $N_b = 21$, $n = 60$, $w = 30$, $\ell = 30$.

$\delta g(x,\chi) = g_0(x,\chi) - g_C(x,\chi)$ given by

$$\delta g(x,\chi) = N_b \left[\frac{N_b - 1}{2} - \chi + \frac{1}{2}\chi x^2 \right] - (1-x)(x + N_b - 1) - 2\chi x(1 - \frac{x}{2}) - \frac{\chi(\chi-1)}{2}(1 - \frac{x^2}{2}) \quad (17)$$

We first study the derivative with respect to ℓ of this difference in the domain $\ell < [(N_b - 1)w/4]$. After simple calculations we obtain that for large N_b and n (we treat here ℓ as a continuous variable) $\mathrm{d}\delta g/\mathrm{d}\ell < 0$ in the domain considered. A similar calculation shows that in the domain $(N_b - 1)/4 < \ell < 2n$, the function $\delta g(\ell, n, \chi)$ is increasing with ℓ (at least for N_b large enough). These results thus show that the minimum of δg is actually reached at the intersection of the two curves and which occurs for

$$\ell_{opt} = \frac{(N_b - 1)w}{4} \quad (18)$$

This expression for ℓ_{opt} is actually independent from the exact form of δg as long as it is decreasing for $\ell < \ell_{opt}$ and increasing above ℓ_{opt} which we verified numerically. We compare the theoretical prediction Eq. (18) with numerical results in Fig. 11, and we see that for N_b large enough (here, typically $N_b > 10$) this prediction is in excellent agreement with data.

We can understand this value of ℓ_{opt} with the following simple argument. If ℓ is small most paths connecting nodes from different branches will go through 0 and we expect $\delta g > 0$. When ℓ is increasing more paths will go through the loop and will increase the value of g_C. However, when ℓ is too large, paths connecting the (large) fraction of nodes located on the lower branches will go through 0 again. In order to get a sufficient condition on ℓ_{opt}, we consider the path between the node C on the branch B_1 and the corresponding node C' on the furthest branch $(N_b - 1)/2$. The optimal value for ℓ_{opt} is then such that the cost of the path from C to C' through 0 and which is 2ℓ is equal to the cost on the loop which is given by $w(N_b - 1)/2$. This immediately gives the result $\ell_{opt} \approx w(N_b - 1)/4$.

The threshold quantity w_c is obtained by imposing that the minimum of $\delta g(\ell = \ell_{opt})$ is equal to zero. Using the approximate form Eq. 17, we can show that

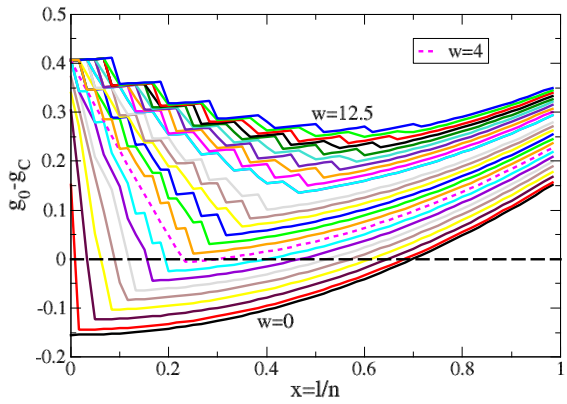


FIG. 10: $\delta g(\ell)$ versus ℓ (for $N_b = 15$ and $n = 60$ here) and for different values of w in the range $[0, 12.5]$. For values less than a threshold ($w_c \approx 4$ shown here by a dotted line) there is a minimum that is negative.

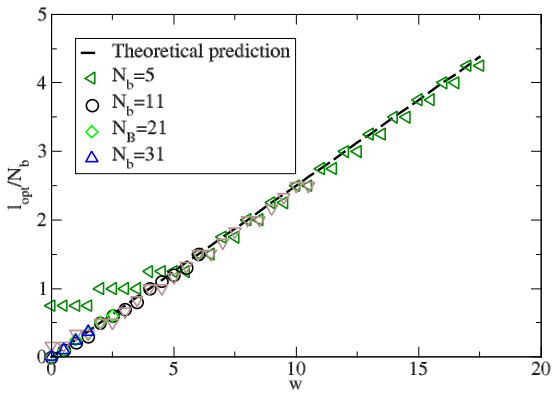


FIG. 11: Comparison between the theoretical prediction (Eq. 18) and numerical results for ℓ_{opt} (for $w < w_c$). For large value of N_b the prediction is excellent. We note that ℓ_{opt} exists for $w < w_c$ and w_c decreases with N_b which implies that the range over which we can see a linear behavior is decreasing as $1/N_b$ (here $n = 40$).

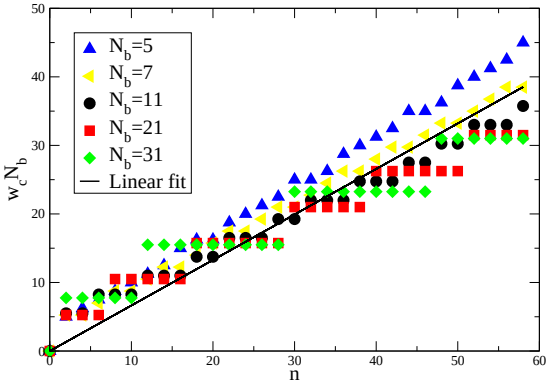


FIG. 12: Value of $w_c N_b$ versus n. The collapse is reasonably good and is in agreement with our theoretical result Eq. 20. We observe plateaus that are due to the discrete values of ℓ and n. The straight line is a linear fit which gives $\kappa_{emp} \approx 0.66$ ($r^2 = 0.96$).

the minimum is actually obtained for $\ell = \ell_{opt}$ and for $\chi = (N_b - 1)/2$. We thus have to consider the quantity $\delta g(\ell_{opt}, n, \chi = (N_b - 1)/2)$ which for large N_b is behaving as large N_b

$$\delta g(\ell_{opt}) \approx \frac{N_b^2}{8}\left[\frac{5}{2}\left(\frac{wN_b}{4n}\right)^2 - 1\right] \tag{19}$$

(details of this calculation are given in appendix) and we therefore obtain

$$w_c \approx \kappa \frac{n}{N_b} \tag{20}$$

where $\kappa = 4\sqrt{\frac{2}{5}}$ in this approximation. We can understand the scaling for w_c with the simple following argument. Indeed, a necessary condition on w is that ℓ_{opt} must be less than n. This gives the condition

$$w < \widetilde{w}_c = 4\frac{n}{N_b} \tag{21}$$

This threshold $\widetilde{w}_c$ is a priori larger than the exact value, as we imposed here a necessary condition, but allows to understand in a simple way the scaling of w_c with n and N_b. We test the scaling for w_c by plotting (Fig. 12) $w_c N_b$ versus n and which should be linear. We indeed observe a reasonable agreement with the linear behavior predicted by our analysis, where the differences are probably due to the small values of N_b used for the numerical calculations. The linear fit however gives a prefactor $\kappa_{emp} \approx 0.66$ which is far from the value obtained within our simple approximation scheme. The important fact is that our approximation is able to predict the correct scaling and it could maybe be possible to find more refined approximations in order to get a better estimate for the prefactor κ. Finally, we note here that w_c is independent from ℓ which can be understood by the fact that w_c gives a condition for the existence of $\ell_{opt} < n$.

Finally, we note that when $w_c > 1$, the case $w = 1$ displays then a negative minimum and we can observe a very central loop. This case is particularly interesting as it corresponds to the 'topological' case for which the distance is the minimum number of hops. This will then happen when there are few branches, or if the branches are large enough.

DISCUSSION

The main purpose of this paper is to shed light on the appearance of non-trivial patterns made of very central nodes (or links) in real-world planar graphs. In particular, we focused on the existence of very central loops that are commonly observed in random planar graphs. We proposed a toy model that shows that indeed a loop at a certain distance from the center can be more central than the physical center itself. The condition for the existence of such a phenomenon is that the weight on the loop has to be small enough and in this toy model we showed the existence of a threshold value w_c. This threshold depends on the size and number of radial branches, highlighting their crucial role. This result allows us to understand the appearance of very central loop even in the topological case where the shortest topological distance is used for computing the BC: if the extension of the network is large compared to the number of radial branches, w_c can be larger than one $w_c > 1$ and central loops for $w = 1$

can be observed. In ordered systems – such as lattices – the effective number of branches is too large leading to a very small w_c and therefore prohibits the appearance of central loops in the 'topological' case ($w = 1$). In real-world planar graphs where randomness is present, the absence of some links can lead to a small number of 'effective' radial branches which in the framework of the toy model implies a large value of w_c and therefore a large probability to observe central loops.

In the case of roads, if we assume that the weight is inversely proportional to the velocity, our result predicts that the velocity on the loop has to be large enough in order to be very central and faster than going through the center. A possible direction of future studies could then to include more precisely different velocities and also to include congestion effects.

More generally, further studies are needed in order to understand the variety of patterns induced by randomness in planar graphs, and we believe that this study is a step towards this direction.

ACKNOWLEDGEMENTS

BL and MB thank Jean-Marc Luck for interesting discussions and valuable suggestions. MB thanks Gourab Ghoshal for stimulating exchanges on this problem.

APPENDIX

BC at the center

Complete formula

The general expression for the BC at the center is

$$g_0(w) = n^2 N_b(N_b-1)/2 - (a_1^0 + a_2^0 + a_3^0) \tag{22}$$

In order to analyze this quantity, we separate the branches into two parts. The first part is composed of nodes at a distance lower than l from the center (the 'lower' part of the branch), and the second part consists in nodes that are at a distance greater or equal than l from the center (the 'upper' part of the branch). When there is no loop, the total number of shortest paths going through the center is (N_{tot} denotes the total number of nodes)

$$\begin{aligned} g_0(w=\infty) &= \binom{N_{tot}-1}{2} - \binom{n}{2} \cdot N_b \\ &= n^2 N_b(N_b-1)/2 \end{aligned} \tag{23}$$

This expression gives the betweenness centrality at the center without the loop, and we can calculate $g_0(w)$ by removing from $g_0(\infty)$ all the shortest paths that go through the loop and not through the center. This can be computed by distinguishing different types of paths: the quantity a_1^0 counts the number of shortest paths going through the loop and connecting nodes both located on the upper part of branches; the quantity a_2^0 counts these paths connecting an upper part to a lower one; and a_3^0 counts the paths connecting nodes both located in lower parts of branches.

We note that due to the symmetry of this network, it is enough to consider paths from a given branch to the others and to multiply at the end of the calculation by the number of branches.

The central point for calculating the centralities $g_0(w)$ and $g_C(w)$ is to compare the length of the path through 0 and the path through the loop. For a node located at distance s on branch 0 and a node at distance t on a branch j (where j goes from 1 to $N_b - 1$), this condition can be written as

$$s + t < |\ell - s| + |\ell - t| + wj \tag{24}$$

If this inequality is satisfied the path will go through 0 and otherwise the loop is more interesting. We thus have to count the pair of nodes that satisfies this inequality and for this we distinguish three different cases: s and t in the lower part of branches, s and t in the lower and upper parts respectively, and finally both s and t in the upper part of branches. In the following we introduce two quantities:

$$X_j = \min\left(\left[\frac{2j}{w}\right], \frac{N_b-1}{2}\right) \tag{25}$$

and

$$\begin{aligned} P_j = &\frac{1}{2}\theta(N_b \text{ odd})\theta(\frac{2j}{w} = X_j) \\ &+ \theta(N_b \text{ even})\left(\frac{2}{3}\theta(\frac{N_b}{2} = \frac{2j}{w}) + \frac{1}{2}\theta(\frac{N_b}{2} < \frac{2j}{w})\right) \end{aligned}$$

where $\theta(condition) = 1$ if condition is true, and is 0 otherwise. This quantity P_j takes into account the fact that the shortest path going through the center or through the loop have the same length. If $2j/w = X_j$ we have to divide by two the number of paths and $X_j - P_j$ is the fraction of paths going through the center. In addition, if N_b is even and $N_b/2 = 2j/w$, we have 3 different paths with the same length that connect two nodes on opposite branches: one path through the center, another 2 paths on each direction of the loop. In this case, there are 2/3 of paths to remove in order to get g_0. in each case $X_j - P_j$ multiplies the total flow.

Calculation of a_1^0 There are $(n - \ell + 1)$ nodes in the upper part of one branch and therefore $N_b(n-\ell+1)^2$ possible pairs between two nodes of two distincts upper parts. The number of branches with nodes that will deviate from the center and will use the loop is given by

X_ℓ. We have however to consider cases where there are shortest paths that equivalently go through the center or via the loop, and this is precisely what is counted by P_ℓ. The coefficient a_1^0 is then given by

$$a_1^0 = N_b(n-\ell+1)^2(X_\ell - P_\ell) \tag{26}$$

We can recover this result by noting that for nodes s and t belonging to the upper part of different branches B_1 and B_j, the condition that the path through the center is longer than on the loop is

$$s + t > s - \ell + t - \ell + wj \tag{27}$$

which leads to $2\ell > wj$ and a number of branches given by $[2\ell/w] = X_\ell$. We have an equality when $2\ell = wj$ and this happen when $X_\ell = 2\ell/w$ and gives a factor $1/2$ in the BC.

Calculation of a_2^0 For the a_2^0 coefficient, only paths between the lower part and the upper part are considered. We consider a node (in the first branch) in the lower part $s < \ell$ and another one in the upper part on $t > \ell$ in the branch j. The path from s to t is deviated from zero if the following condition is met

$$\begin{aligned} &s + t > |\ell - s| + |\ell - t| + wj \\ &\Rightarrow 2s > wj \end{aligned} \tag{28}$$

which means that the number of such paths is given by the number of nodes in the upper part $(n-\ell+1)$ times the number of branches that satisfy this condition: $j < \left[\frac{2s}{w}\right]$. We have to sum over $s \in [\![1, \ell]\!]$ and to multiply by $2N_b$ which takes into account both paths (from the upper to the lower and from the lower to the upper part of branches) and obtain

$$a_2^0 = 2N_b(n-\ell+1)\sum_{s=1}^{\ell-1}(X_s - P_s) \tag{29}$$

(the term P_s takes into account the degeneracies of paths).

Calculation of a_3^0 The quantity a_3^0 represents deviation between pairs of nodes both located in the lower part of branches. in this case, the condition on paths is

$$s + t > \ell - s + \ell - t + wj \tag{30}$$

which implies that the branches where we have a deviation from 0 are such that

$$j < \frac{2(s+t-\ell)}{w} \tag{31}$$

and we then obtain

$$a_3^0 = N_b \sum_{s=1}^{\ell-1}\sum_{t=1}^{s-1}(X_{s+t-\ell} - P_{s+t-\ell}) \tag{32}$$

We therefore finally obtain the exact expression

$$g_0(w) \approx n^2 N_b(N_b-1)/2 - (a_1 + a_2 + a_3) \tag{33}$$

with

$$a_1 = N_b(n-\ell+1)^2 \cdot (X_\ell - P_\ell) \tag{34}$$

$$a_2 = 2N_b(n-\ell+1)\sum_{s=1}^{\ell-1}(X_s - P_s) \tag{35}$$

$$a_3 = N_b \sum_{s=1}^{\ell-1}\sum_{t=1}^{s-1}(X_{s+t-\ell} - P_{s+t-\ell}) \tag{36}$$

where $X_j = \min(\left[\frac{2j}{w}\right], \frac{N_b-1}{2})$ and P_j given above. The sums entering these expressions are however difficult to handle and we therefore resort to approximations that are detailed in the main text.

Simplification

Another way to recover the approximation discussed in the main text is to neglect small deviation terms (ie. to impose $P_j = 0$), and we then obtain

$$g_0(w) \approx n^2 N_b(N_b-1)/2 - (a_1 + a_2 + a_3) \tag{37}$$

with

$$a_1 = N_b(n-\ell+1)^2 \cdot X_\ell \tag{38}$$

$$a_2 = 2N_b(n-\ell+1)\sum_{s=1}^{\ell-1} X_s \tag{39}$$

$$a_3 = N_b \sum_{s=1}^{\ell-1}\sum_{t=1}^{s-1} X_{s+t-\ell} \tag{40}$$

We are still left with the sums to compute and the simplest approximation we can think of (and that can be used for g_C too) is to choose $X_j \approx X_\ell = \chi$ leading to

$$a_1 \approx N_b(n-\ell+1)^2\chi \tag{41}$$

$$a_2 \approx 2N_b(n-\ell+1)\chi(\ell-1) \tag{42}$$

$$a_3 \approx N_b\chi\frac{(\ell-2)(\ell-1)}{2} \tag{43}$$

leading to the result Eq. 11 (in the limit $n, \ell \to \infty$ with $x = \ell/n$ fixed). This is obviously a very crude approximation and it could certainly be refined in order to get a more accurate expression for g_0. However, as we will see in the next section, the expression for g_C is much more involved and we need an approximation scheme that can be applied to both quantities g_0 and g_C, which seems to be a difficult task that we leave for future studies.

BC for the loop

Complete formula

The first term $g_C(w=\infty) = (n-\ell)(n(N_b-1)+\ell)$ is the number of shortest path passing through C when there is no loop. The quantity $(n-\ell)$ is the number of nodes in the upper part of the branch B_1 and the number of node in the rest of the network is $(n(N_b-1)+\ell)$. Similarly as for g_0, we start from the quantity computed for $w=\infty$ and add the number of paths that will go through the loop and obtain

$$g_C(w) = (n-\ell)(n(N_b-1)+\ell) + (a_1^C + a_2^C + a_3^C) \quad (44)$$

The first term is considering only deviation from the upper part of the branches to other upper parts. For all other branches $k = 2, \ldots, N_b/2$ (by symmetry and for N_b even), if $wk < 2\ell$, there are $((k-1)*(n-\ell+1)*(n-\ell+1))$ additional shortest paths going via the loop. By summing over k and taking into account multiple paths, we obtain

$$a_1^C = (n-\ell+1)^2\left(\frac{1}{2}(X_{\ell-w/2}+1)X_{\ell-w/2} - P_\ell^1\right) \quad (45)$$

The coefficient a_2^C counts the paths from the upper part of a branch to the lower part of another branch. This path will go through the loop if $wi < 2j$ with i going from 1 to $N_b/2$ and j from 1 to $\ell-1$. We then obtain

$$a_2^C = 2(n-\ell+1)\sum_{j=1}^{\ell-1}\left(\frac{X_j(X_j+1)}{2} - P_j^2\right) \quad (46)$$

Finally, the coefficient a_3^C corresponds to additional shortest paths from lower part to lower part and going through the loop. When $wi < 2(j+k-\ell)$, with $i = 1, \ldots, N_b/2$, $j = 1, \ldots, \ell-1$, and k running from $\ell-j+1$ to $\ell-1$, there are $(i+1)$ new shortest paths added at point C. Summing over i, we then obtain

$$a_3^C = \sum_{j=1}^{\ell-1}\sum_{k=1}^{j-1}\left(\frac{X_{j+k-\ell}(X_{j+k-\ell}+3)}{2} - P_{j+k-\ell}^3\right) \quad (47)$$

The quantities $P_j^i,\ i = 1, ..., 3$ correspond to the correction needed when the path going through the loop has the same weight as the path going through 0.

For the part 1:

$$\begin{aligned} P_j^1 = &\frac{X_j}{2}\theta(\frac{2j}{w} = X_j)\theta(\frac{N_b}{2} \neq X_j) \\ &+ \frac{N_b/2-1}{3}\theta(\frac{N_b/2}{2} = \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right] \\ &+ \frac{N_b/2-1}{2}\theta(\frac{N_b}{2} < \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right]) \end{aligned} \quad (48)$$

For the part 2:

$$\begin{aligned} P_j^2 = &\frac{X_j}{2}\theta(\frac{2j}{w} = x_j)\theta(\frac{N_b}{2} \neq X_j) \\ &+ \frac{2(N_b/2)}{3}\theta(\frac{nb}{2} = \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right]) \\ &+ \frac{N_b/2}{2}\theta(\frac{N_b}{2} < \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right]) \end{aligned} \quad (49)$$

For the part 3:

$$\begin{aligned} P_j^3 = &\frac{X_j+1}{2}\theta(\frac{2j}{w} = X_j)\theta(\frac{N_b}{2} \neq X_j) \\ &+ \frac{2X_j+2}{3}\theta(\frac{N_b}{2} = \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right]) \\ &+ \frac{X_j}{2}\theta(\frac{N_b}{2} < \frac{2j}{w})\theta(\frac{N_b}{2} = \left[\frac{N_b}{2}\right]) \end{aligned} \quad (50)$$

Simplification

Here also, we can recover the approximate formula (Eq. 15) by using the same approximation as described above for g_0: we neglect small deviation terms ($P_j = 0$) and we assume that $X_j = X_\ell \approx \chi$ for all j which gives

$$g_C(w) = (n-\ell)(n(N_b-1)+\ell) + (a_1^C + a_2^C + a_3^C) \quad (51)$$

with

$$\begin{aligned} a_1^C &= (n-\ell+1)^2\frac{1}{2}(X_{\ell-w/2}+1)X_{\ell-w/2} \\ &\approx (n-\ell+1)^2\frac{\chi(\chi-1)}{2} \end{aligned} \quad (52)$$

$$\begin{aligned} a_2^C &= 2(n-\ell+1)\sum_{j=1}^{\ell-1}\frac{X_j(X_j+1)}{2} \\ &\approx 2(n-\ell+1)(\ell-1)\frac{\chi(\chi+1)}{2} \end{aligned} \quad (53)$$

$$\begin{aligned} a_3^C &= \sum_{j=1}^{\ell-1}\sum_{k=1}^{j-1}\frac{X_{j+k-\ell}(X_{j+k-\ell}+3)}{2} \\ &\approx \frac{\chi(\chi+3)}{2}\ \frac{(\ell-2)(\ell-1)}{2} \end{aligned} \quad (54)$$

where $\chi = \min(\left[\frac{2j}{w}\right], \frac{N_b-1}{2})$. Summing all these terms and taking the limit $n, \ell \to \infty$ with $\ell/n = x$ fixed, we recover the approximation Eq.15.

Calculation of w_c

We start with the expression Eq. 17

$$\delta g(x,w) = N_b\left[\frac{N_b-1}{2} - \chi + \frac{1}{2}\chi x^2\right] - (1-x)(x+N_b-1) - 2\chi x(1-\frac{x}{2}) - \frac{\chi(\chi-1)}{2}(1-\frac{x^2}{2}) \quad (55)$$

where $\chi = \frac{N_b-1}{2}$. In the limit $N_b \gg (x-1)$ we obtain (keeping terms growing with N_b)

$$\delta g(x,w) \simeq N_b\left[\frac{N_b-1}{4}x^2\right] - (1-x)N_b - 2\frac{N_b-1}{2}x(1-\frac{x}{2}) - \frac{N_b-1}{4}(\frac{N_b-1}{2}-1)(1-\frac{x^2}{2})$$

which behaves at leading order as

$$\delta g(x,w) \simeq \frac{N_b^2}{8}\left(\frac{5}{2}x^2 - 1\right) \quad (56)$$

The minimum is crossing zero at

$$x^2 = \frac{2}{5} \quad (57)$$

which implies for $\ell = \ell_{opt} \simeq N_b w/4$

$$w_c \simeq \sqrt{\frac{2}{5}}\frac{4n}{N_b} \quad (58)$$

* Electronic address: marc.barthelemy@cea.fr

[1] Clark, J., Holton, D. A. A first look at graph theory (Vol. 1). Teaneck, NJ: World Scientific (1991).

[2] Tutte, W.T. A census of planar maps. *Canad. J. Math* **15**, 249-271 (1963).

[3] Bouttier, J., Di Francesco, P. & Guitter, E. Planar maps as labeled mobiles. *Electron. J. Combin* **11**, R69 (2004).

[4] Ambjorn, J. & Jonsson, P. *Quantum geometry: a statistical field theory approach* (Cambridge University Press, Cambridge, UK, 1997).

[5] Mileyko, Y., Edelsbrunner, H., Price, C.A. & Weitz J.S. Hierarchical ordering of reticular networks. *PLoS One* **7**, e36715 (2012).

[6] Katifori, E. & Magnasco, M.O. Quantifying loopy network architectures. *PLoS One* **7**, e37994 (2012).

[7] Katifori, E., Szollosi GJ, Magnasco, M.O. Damage and fluctuations induce loops in optimal transport networks. *Physical Review Letters* 104:048704 (2010).

[8] Hu, D., Cai, D. Adaptation and optimization of biological transport networks. *Physical review letters* 111:138701 (2013).

[9] Barthelemy, M. Spatial Networks. *Phys. Rep.* **499**, 1-101 (2011).

[10] Haggett, P. & Chorley, R.J. 1969 *Network analysis in geography*. London: Edward Arnold.

[11] Hillier, B., Hanson, J. The social logic of space (Vol. 1). Cambridge: Cambridge university press (1984).

[12] Marshall, S. 2006 *Streets and Patterns*. Abingdon: Spon Press.

[13] B. Jiang & C. Claramunt 2004 Topological analysis of urban street networks. *Environment and Planning B : Planning and design* **31**, 151-162.

[14] Roswall, M., Trusina, A., Minnhagen, P. & Sneppen, K. 2005 Networks and cities: an information perspective *Phys. Rev. Lett.* 94, 028701.

[15] Porta, S., Crucitti, P. & Latora, V. 2006 The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design* **33**, 705–72.

[16] Porta, S., Crucitti, P. & Latora, V. 2006 The network analysis of urban streets: a dual approach *Physica A* 369, 853866.

[17] Lammer, S., Gehlsen, B. & Helbing, D. 2006 Scaling laws in the spatial structure of urban road networks. *Physica A* **363**, 89–95.

[18] Crucitti, P., Latora, V. & Porta, S. 2006 Centrality measures in spatial networks of urban streets. *Phy. Rev. E* **73**, 0361251-5.

[19] Cardillo, A., Scellato, S., Latora, V. & Porta, S. 2006 Structural properties of planar graphs of urban street patterns. *Phys. Rev. E* **73**, 066107.

[20] Xie, F. & Levinson, D. 2007 Measuring the structure of road networks. *Geographical Analysis* **39,** 336–356.

[21] Jiang, B. 2007 A topological pattern of urban street networks: universality and peculiarity *Physica A* 384, 647655.

[22] Masucci, A. P., Smith, D., Crooks, A. & Batty, M. 2009 Random planar graphs and the London street network *Eur. Phys. J. B* 71, 259271.

[23] Chan, S. H. Y., Donner, R. V. & Lammer, S. 2011 Urban road networks- spatial networks with universal geometric features? *Eur. Phys. J. B* 84, 563577.

[24] Courtat, T., Gloaguen, C. & Douady, S. 2011 Mathematics and morphogenesis of cities: A geometrical approach. *Phys. Rev. E* **83**, 036106.

[25] Strano, E., Nicosia, V., Latora, V., Porta, S. & Barthelemy, M. 2012 Elementary processes governing the evolution of road networks. *Nature Scientific Reports* **2**:296.

[26] Barthelemy, M., Bordin, P., Berestycki, H. & Gribaudi, M. Self-organization versus top-down planning in the evolution of a city. *Nature Scientific Reports* **3**:2153 (2013).

[27] Viana, M. P., Strano, E., Bordin, P., Barthelemy, M. The simplicity of planar networks. *Scientific reports*, 3 (2013).

[28] Louf, R., Barthelemy, M. A typology of street patterns. Journal of The Royal Society Interface, 11(101), 20140924 (2014).

[29] Porta, S., Romice, O., Maxwell, J. A., Russell, P., & Baird, D. 2014 Alterations in scale: Patterns of change in main street networks across time and space *Urban Studies* 0042098013519833.

[30] Aldous, D.J., Shun, J. Connected Spatial Networks over Random Points and a Route-Length Statistic. *Stat. Sci.* **25**, 275-288 (2010).

[31] L.C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* **40**, 35–41 (1977).

[32] Ashton, D. J., Jarrett, T. C., Johnson, N. F. Effect of congestion costs on shortest paths through complex networks. *Physical review letters* 94:058701 (2005).

[33] Jarrett, T. C., Ashton, D. J., Fricker, M., Johnson, N. F. Interplay between function and structure in complex networks. *Physical Review E*, 74:026116 (2006).

[34] Jungnickel, D. Graphs, networks and algorithms. Berlin: Springer, 2008.

[35] Brandes, U.. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* 25:163-177 (2001).

[36] M. Barthelemy, Betweenness Centrality in Large Complex Networks, *Eur. Phys. J. B.* **38**:163 (2004).

[37] Mapzen Weekly OSM Metro Extracts: `https://mapzen.com/metro-extracts`

[38] Rheinwalt, A., Marwan, N., Werner, P., Gerstengarbe, F. W. Boundary effects in network measures of spatially embedded networks. *Europhys. Lett.*, 100(2), 28002 (2012).

[39] Gil, J. Street network analysis edge effects: Examining the sensitivity of centrality measures to boundary conditions. Environment and Planning B: Planning and Design, 0265813516650678 (2016).