Node diversification in complex networks by decentralized colouring

Richard Garcia-Lebron
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

David J. Myers
Air Force Research Laboratory, Rome, NY, USA

Shouhuai Xu
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

and

Jie Sun†
Department of Mathematics, Clarkson University, Potsdam, NY, USA
and Department of Mechanical & Aerospace Engineering, Clarkson University, Potsdam, NY, USA
†Corresponding author. Email: sunj@clarkson.edu

Edited by: Ernesto Estrada

[Received on 23 April 2018; editorial decision on 7 November 2018; accepted on 10 November 2018]

We develop a decentralized colouring approach to diversify the nodes in a complex network. The key is the introduction of a local conflict index (LCI) that measures the colour conflicts arising at each node which can be efficiently computed using only local information. We demonstrate via both synthetic and real-world networks that the proposed approach significantly outperforms random colouring as measured by the size of the largest colour-induced connected component. Interestingly, for scale-free networks further improvement of diversity can be achieved by tuning a degree-biasing weighting parameter in the LCI.

Keywords: decentralized colouring; network connectivity; iterative optimization.

1. Introduction

Complex networks have been serving as structural models for many complex systems in the real world, including social networks as well as engineering and biological systems [1]. Among the many properties of complex networks, connectivity and its determining factors have received considerable attention in recent years [2, 3], especially the role played by connectivity in dynamical processes such as synchronization [4, 5], information communication [6] and epidemic spreading and control [7, 8]. Enhanced connectivity is desirable from the point of view of network functioning, such as communication, information sharing and transfer of energy. However, connectivity can be a double-edged sword as it is intimately related to the spreading of virus and disease [7–15]. For computer networks, including the peer-to-peer networks that serve as the networking infrastructure of the widely used blockchains in general and cryptocurrency in particular [16], the adverse effect is rooted in the monoculture of computer software systems: when all computers run the same software operating system (e.g., Microsoft Windows), a single software vulnerability can cause many computers to be compromised especially when the underlying network is highly connected [17, 18]. For cyber security applications, it has been advocated to overcome this...
monoculture problem by introducing software diversification \cite{19, 20}. For complex networks representing human social contacts, coping with the spreading of epidemic diseases demands the use of vaccines, which are often scarce especially when the diseases are new variants of viruses, to immunize few people in the network while maximize the suppression effect, which is fundamentally related to the problem of disrupting network clusters. The concept of diversification is also important in socioeconomics, given a strong link between structural network diversity and economic development \cite{21}.

In this article, we focus on a particular type of network diversification problem, which is closely related to the problem of graph colouring. A network of coupled units can be abstracted as a graph $G = (V, E)$ where a pair of nodes $(u, v) \in E$ if and only if $u$ and $v$ are connected by an edge in the graph. A colouring of the graph is an assignment function from a set $V$ of nodes to a set of $q$ colours,

$$C : V \to \{1, 2, \ldots, q\}, \quad (1.1)$$

where $C(u)$ denotes the colour assigned to node $u$. Upon colouring, an edge $(u, v) \in E$ is called defective if $C(u) = C(v)$, that is, if nodes $u$ and $v$ have the same colour. Unlike non-defective edges, the defective edges are assumed to be the transmission channels for the spread of undesired information such as computer virus. In addition to the problem of colouring planar graphs (also known as the “map colouring problem” for which the four colour theorem is about), the graph colouring abstraction has several applications, for example in task scheduling problems (including computer register allocation) \cite{22}, frequency assignment and planning in wireless communication systems \cite{23}, and more recently in studying the virus propagation problems in various computer networks \cite{20, 24–28}. In these applications, the goal of colouring is to diversify the colours of the nodes so that the network is ‘disrupted’ into as small as possible independent sets—sets of isolated components made up of defective edges.

In this article, we develop a decentralized colouring approach based on iterative minimization of a local conflict index (LCI), which is a quantity that is directly computable from the local information at each node. We validate the effectiveness of our method for both random and real-world networks, and found that for scale-free (SF) networks further improvement can be achieved by using a degree-biasing weighting scheme.

2. Decentralized network colouring

For graph $G$ and colour assignment $C$, we define the LCI at node $u$ as

$$\text{LCI}(u) = \sum_{v \in \mathcal{N}_u} w_v \delta(C(u), C(v)), \quad (2.1)$$

where $\mathcal{N}_u$ denotes the set of neighbours of $u$ in $G$, and $\delta(i, j) = 1$ if $i = j$ and 0 otherwise. The weight $w_v$ can be used to adjust the relative contribution of a defective edge $(u, v)$ to LCI$(u)$.

As a first attempt at exploring the effect of the weight, we consider weights of the form

$$w_v = k_v^\beta, \quad (2.2)$$

where $k_v = |\mathcal{N}_v|$ denotes the degree of node $v$. This weighting scheme allows one to tune the relative influence of node degree on LCI by adjusting exponent $\beta$. For the special case $\beta = 0$, LCI$(u)$ gives the number of defective edges at node $u$ since all weights $w_v = 1$. This is the scenario considered in classical statistical physics such as the Potts model \cite{29}. For any node with degree $k > 1$: when $\beta \to -\infty$, LCI $\to 0$ regardless of the particular colour assignment; on the other hand, when $\beta \to +\infty$, the value
Fig. 1. Minimization of LCI at a given node $u$ depends on the choice of $\beta$ in Equation (2.1). As examples: (a) when $\beta = 0$, LCI($u$) is minimized by choosing a colour for $u$ that results in the smallest number of defective edges that connect to $u$. (b) When $\beta = 1$, LCI($u$) is minimized by choosing a colour for $u$ so that the total degree of the defective neighbours of $u$ is minimized.

of LCI($u$) is dominated by the maximum degree of the neighbours of $u$ that have the same colour as $u$’s. When $\beta > 0$ ($\beta < 0$) and proper colouring at node $u$ is not possible, the minimization of LCI($u$) is generally achieved by biasing the colouring of $u$ toward the low-degree (high-degree) neighbours of $u$ that have the same colour as $u$’s. An illustration is provided in Fig. 1, which visually highlights the following: when $\beta = 0$, LCI($u$) is minimized by choosing a colour that results in the least number of defective edges that connect to $u$; when $\beta = 1$, LCI($u$) is minimized when the total degree of the defective neighbours of $u$ (i.e., the neighbours connected to $u$ over defective edges) is minimized. A key feature of LCI is that it can be computed without global knowledge of the network. Importantly, the determination of LCI($u$) only involves information about the colours as well as the degrees of $u$ and its neighbours.

Given a network $G$, a fixed number of colours $q$ and the choice of weights $w_v$, the iterative minimization of LCI over the nodes of a network gives rise to a dynamic colouring process, which we call dynamic decentralized colouring (DDC). Initially, each node is assigned a colour chosen uniformly at random from the $q$ colours (nodes with the same colours form a site percolation graph with ‘occupation’ probability $1/q$ [30]). Then, in each iteration, a node $u$ is selected randomly, and its colour will be updated to minimize LCI($u$) given information about the colours and degrees of the neighbours of $u$. When there is no defective edge associated to node $u$, the colour of $u$ remains unchanged. On the other hand, when there is at least one defective edge, $u$ updates its colour to minimize LCI($u$). When multiple colour choices yield the same minimal LCI($u$), the new colour of $u$ will be chosen randomly among these minimizers. Empirically we found the algorithm converges in $O(ng(n))$ iterations in terms of the fraction of defection edges, that is, $O(g(n))$ colour updates per node. The function $g(n)$ generally decreases as the number of colours $q$ increases, from linear (when $q < \chi(G)$) to logarithmic (when $q \gtrsim \chi(G)$) for various types of networks.

3. Results

3.1 Measures of diversity

In order to quantify the network diversity upon colouring, we consider the following two measures. One measure is the fraction of defective edges, given by

$$f_d = \frac{\text{number of defective edges}}{\text{total number of edges}}.$$  

(3.1)
The case of $f_d \to 0$ occurs when there is no defective edge whereas $f_d \to 1$ if and only if all edges are defective, which, for a connected network, can only occur if all nodes have the same colour. In general $0 \leq f_d \leq 1$, with a smaller value corresponding to a better diversity. However, $f_d$, as a local measure, has some undesirable limitations. For example, the colouring may yield a large colour-induced component even if $f_d \approx 0$. We therefore also consider a global diversity measure

$$R_{\text{max}} = \text{size of the largest colour-induced component}. \quad (3.2)$$

For a given colouring, $R_{\text{max}}$ can be interpreted as the maximum range of spread as it is the maximum number of nodes that can be reached from a single node through defective edges.

### 3.2 Random networks

We first explore the global effect of the parameter $\beta$ on the colouring of networks. We consider two networks with the same number of nodes ($n = 1,000$). The first network is generated by the classical Erdős–Rényi (ER) model, where an edge is created between every pair of nodes $(u, v)$ with probability $p$ [31]. Here we choose $p = 0.015$, resulting in a sparse network with average degree $\langle k \rangle \approx np = 15$. The second network is a SF network generated by the configuration model [32, 33] with expected degree distribution $P(k) \approx k^{-\gamma}$. Here we set the degree exponent $\gamma = 2.5$ and minimal degree $k_{\text{min}} = 5$, obtaining a network with average degree $\langle k \rangle \approx 12$. Both networks have a single connected component. Upon colouring, the networks are expected to be disrupted into same-colour components as the number of available colours increases. Figure 2 shows that for the ER network with a fixed number of colours,
both the fraction of defective edges \( f_d \) (Fig. 2(a,b)) and the maximum range of damage \( R_{\text{max}} \) (Fig. 2(c,d)) are minimized when \( \beta \approx 0 \). Interestingly, the same does not hold true for the SF network. The optimal value of \( \beta \) for the SF network in fact depends on the number of colours. When there are very few colours available, the optimal \( \beta \approx 0 \) (Fig. 2(a,c)); on the other hand, with increasing number of colours, the optimal value of \( \beta \gg 0 \) (Fig. 2(b,d)). The fact that \( f_d \) (the fraction of defective edges in the network) is minimized for a non-zero \( \beta \) suggests a non-trivial global effect of the minimization of LCI, because locally the LCI value indeed equals the number of defective edges with the choice of \( \beta = 0 \) (equivalent to greedily minimizing \( f_d \) in each iteration).

Next, we study how the number of colours affects network diversity. We consider both the ER network and the SF network as described in Fig. 2 and compute the fraction of defective edges \( f_d \) as well as the maximum range of damage \( R_{\text{max}} \) as functions of the number of colours. Results from three distinct algorithms are compared: (1) random colouring by choosing a colour for each node uniformly at random from the \( q \) colours; (2) iterative LCI minimization with \( \beta = 0 \); and (3) iterative LCI minimization with \( \beta = \beta^* \), which is the optimal value of \( \beta \) for the given number of colours found numerically by searching over \( \beta \in (-2, 2) \). The results are shown in Fig. 3. In comparison to random colouring, minimization of LCI leads to substantially faster decay of both \( f_d \) and \( R_{\text{max}} \) as function of \( q \), both of which reach the saturation level with \( q = 10 \) colours, at which point a proper colouring is essentially achieved. For the ER network, there is no significant difference whether \( \beta \) is chosen to be 0 or the actual optimal value (which can in fact differ from \( \beta \)), suggesting that \( \beta = 0 \) effectively leads to optimal diversity independent of the number of colours \( q \). This is consistent with the results shown in Fig. 2(a,c) for ER network. However, for the SF network, it is in fact possible to achieve significant improvement of network diversity by optimally choosing \( \beta \), and such choice is not universal as the ER case but rather depends on the network structure as well as the number of colours available.

![Fig. 3](https://academic.oup.com/comnet/article-abstract/7/4/554/5229540/16 October 2019)

**Fig. 3.** Dependence of network diversity on the number of colours in the three colouring schemes: randomly selecting a colour for each node (random), iterative minimization of the LCI [see Equation (2.1)] using \( \beta = 0 \) and the optimal value of \( \beta \). (a) Fraction of defective edges \( f_d \) [as defined in Equation (3.1)] as functions of the number of colours for ER network with \( n = 1,000 \) nodes and edge probability \( p = 0.015 \). (b) Same as (a), for SF network with \( n = 1,000 \) nodes and degree distribution \( P(k) \sim k^{-2.5} \) with minimal degree \( k_{\text{min}} = 5 \). (c,d) Same as (a,b) by considering the maximum range of damage \( R_{\text{max}} \) given by Equation (3.2). In all panels, each data point represents the average over 150 independent runs under the same parameters.
3.3 Networks with two communities

The types of networks we considered so far are unstructured, with no community or other structures. Are networks with community structures easier or harder to diversify? As a first attempt to address this question, we consider a random network model with \( n \) nodes that are split into two groups. Within each group, two nodes are connected with probability \( p_{\text{in}} \); two nodes from different groups are connected with probability \( p_{\text{out}} \). That is, \( p_{\text{in}} \) and \( p_{\text{out}} \) are the within-group (or within-community) and cross-group (cross-community) connection probabilities, respectively. The larger the value of \( p_{\text{in}} \) when compared with \( p_{\text{out}} \), the stronger the community structure, whereas in the case of \( p_{\text{in}} = p_{\text{out}} \) the model reduces to a standard ER model with no community structure. Through numerical simulations, we found that for networks generated by this model, the outcome of decentralized colouring does not depend much on the parameter \( \beta \), similar to the case of ER networks. However, diversity does depend on how strong the communities are, as controlled by the parameters \( p_{\text{in}} \) and \( p_{\text{out}} \). Generally speaking, for networks with similar total numbers of edges, networks with a smaller within-community connection probability \( p_{\text{in}} \) seem easier to diversify (see Fig. 4). This also suggests that the presence of community structures makes it harder to diversify the network.

3.4 Application: decentralized colouring of an email communication network

Finally, noting that email communications represent a primary source of computer virus spreading and the difficulty of managing the system in a centralized fashion, we test our decentralized colouring approach on a real-world email network. The network we consider was constructed from email communications between members of the University Rovira i Virgili [34]. We focus on the largest connected component of the network which contains \( n = 1,133 \) nodes and \( m = 5,451 \) edges. The average degree \( \langle k \rangle = 2m/n \approx 9.62 \) and the maximum degree \( k_{\text{max}} = 71 \), with the degree distribution reasonably resembled by an exponential, \( P(k) \propto \exp(-k/k^*) \) with \( k^* \approx 9.2 \) [34]. In Fig. 5(a) we plot \( R_{\text{max}} \) as functions of \( \beta \) that result from colouring the network via the proposed decentralized algorithm, for varying numbers of colours \( q \). As the number of colours increases past \( q = 4 \), the value of \( R_{\text{max}} \) tends to be the smallest when \( \beta \approx 0 \). As shown in Fig. 5(b), the colouring obtained by the proposed method achieves much better network diversification compared to random colouring, and there is a wide range of number of

![Fig. 4. Impact of community structure on colouring. We apply DDC to random networks with \( n = 1,000 \) nodes for several different number of colours, \( q \). We found that the parameter \( \beta \) has the minimal impact on the outcome, and thus fix \( \beta = 0 \) throughout the experiments. For each network, the nodes are divided into two equal-size groups (communities): nodes within the same community are connected with probability \( p_{\text{in}} \), and nodes from different communities are connected with probability \( p_{\text{out}} \). To compare networks with different strength of community structure, we vary \( p_{\text{in}} \) (within-community connection probability) while fixing \( p_{\text{in}} + p_{\text{out}} = 0.02 \) so that the average degree \( \langle k \rangle \approx \frac{2}{3}(p_{\text{in}} + p_{\text{out}}) = 10 \). We found that network diversification, as measured by \( f_d \) and \( R_{\text{max}} \), are both worsened as \( p_{\text{in}} \) increases, that is, as community structure becomes more pronounced.](https://academic.oup.com/comnet/article-abstract/7/4/554/5229540)
Fig. 5. Dependence of network diversity as measured by $R_{\text{max}}$ as functions of (a) the LCI exponent $\beta$ used in the proposed decentralized colouring under different numbers of colours $q$; and (b) the number of colours under random colouring (‘rand’) and the proposed decentralized colouring with $\beta = 0$ and $\beta = 1$, respectively.

colours for which the choice of $\beta = 1$ outperforms the choice of $\beta = 0$. Given that $R_{\text{max}}$ measures the number of computers that can potentially be infected due to (direct and indirect) communications with a single initially infected computer, these results suggest that our decentralized colouring approach can significantly reduce the risk of large-scale virus outbreak. Such prevention is even more effective by optimizing the weight parameter $\beta$ although such optimization depends intricately on both the network structure and resources available (number of ‘colours’).

4. Conclusion and discussion

In this article, we develop a decentralized colouring approach to diversify the nodes of a large complex network. The proposed approach, reminiscent of a self-organized process that involves no central controlling, is achieved by iterative minimization of LCI over the nodes in a network. In our LCI formulation, the case of weighing parameter $\beta = 0$ is equivalent to the antiferromagnetic Potts model at zero temperature [29]. By allowing $\beta > 0$ which tends to avoid defective edges connecting to high-degree neighbours, our approach can achieve a significantly improved diversity so long as there is a relatively abundance of colours. Interestingly, such improvement is only observed in SF networks but not for ER networks which have a rather flat degree distribution. From a theoretical perspective, it remains a challenge to uncover the mechanisms that underlie this discrepancy, which we hope to address in future work. In addition, our colouring algorithm only takes into account the degree of nodes and ignores the effect of clustering. Although the problem is already quite complex and shows rich phenomena even at this relatively simple setting, further incorporation of local clustering information could prove useful for certain types of networks. Furthermore, the preliminary results we reported regarding the dependence of colouring outcome on community structure show that (with fixed number of nodes and edges) networks with more profound community structure are harder to diversify. Interestingly, the graph complement of such networks have recently been reported to promote diffusion and synchronizability [35]. Future research on the relationship between colouring and structural properties of networks will likely yield new insights into how to diversify networks that are more realistic than the rudimentary network models considered here.

Among applications, we note that distributed defective colouring algorithms have often been used as an intermediate step for attaining proper colouring [24, 36]. For a network with maximum degree $k_{\text{max}}$, the state-of-the-art distributed defective colouring algorithm [37] produces colouring-induced subgraphs whose maximum degree is upper bounded by $k_{\text{max}}/q$. Empirically we found our algorithm to typically yield a much smaller maximum degree. This hints future research to design decentralized proper-colouring
algorithm. For cybersecurity applications, we tested our colouring approach on an empirical email communication network and found that the network can be effectively ‘disrupted’ into small disconnected communication components using just a handful of colours. This suggests an opportunity to potentially enhance network communication security via the design and implementation of decentralized software diversification.

Finally, at a fundamental level, the DDC introduced in this article essentially defines a stochastic process on the network. The colouring of networks is intimately related to the asymptotic behaviour of such process. Does the process converge, (if so) does it always converge to the same distribution, how fast, on what types of networks, using how many colours? It is our hope that future work will address these interesting and relevant questions.

Acknowledgments

We thank Daniel ben-Avraham for comments and discussions.

Funding

Simons Foundation (318812 to J.S.); the Army Research Office (W911NF-12-1-0276 to J.S.; W911NF-17-1-0566 to R.G. and S.X.); and NSF (1736209 to R.G. and S.X.). This paper was approved for public release; distribution unlimited: 88ABW-2016-0517.

REFERENCES


