We are particularly interested in quantifying the likelihood of an AS to lie on paths between two other ASes, also known as AS centrality (§3). AS hegemony is a fundamental metric that represents the fraction of paths between any two ASes that pass through a given AS. Significant changes in AS centrality are strong evidences of structural routing changes that might be undesirable. In the literature, AS centrality is commonly measured using Betweenness Centrality (BC). For example, BC enabled researchers to monitor critical ASes at country-level [5], detect disruptive events [2], and select targets for control plane attacks [4]. However, in this paper, we report fundamental shortcomings of BC when used with BGP data (§2), consequently we propose a new centrality metric called AS hegemony (§3).

2 BETWEENNESS CENTRALITY

BC is a fundamental metric that represents the fraction of paths that go through a node. Intuitively one expects high BC scores for transit ASes as they occur on numerous AS paths, and low BC scores for stub ASes. Formally, for a graph G = (V,E) composed of a set of nodes V and edges E, the betweenness centrality is defined as:

\[
BC(v) = \frac{1}{S} \sum_{u,w \in V} \sigma_{u,w}(v)
\]

where \( \sigma_{u,w}(v) \) is the number of paths from u to w passing through v, and S is the total number of paths. BC ranges in [0, 1], but the relative magnitudes of the scores are usually more significant than the absolute values.

Shortcomings with BGP Data: In theory, to compute BC one needs to know the set of all paths in the graph. With BGP data, however, we are restricted to paths bounded to a small number of viewpoints. We found that this singular type of path sampling greatly impairs BC results. To illustrate this, we present an example in Figure 1 with 13 ASes and three viewpoints. If we had viewpoints in all ASes, thus access to all paths in the graph, we would obtain the highest BC score for the transit ISP (.62) and lowest scores for the stub ASes (.15). But, using only paths bound to the three viewpoints, the computed BC scores are substantially different (Sampled BC in Fig.1). Since about a third of the paths converge to each viewpoint, BC values for ASes close to the viewpoints are undesirably high making these ASes look more central than others. This bias is so pronounced that the BC for stub ASes accommodating viewpoints (.38) is twice higher than the BC of one of the regional ISP (.16). Although theoretical studies have already reported that BC is significantly altered by sampling methods [1], this issue has been rarely acknowledged in the networking literature. Mahadevan et al. [3] have reported that BC is not a measure of centrality when computed with network data, but we stress that this issue comes from the non-random, and opportunistic, sampling method used to collect BGP data rather than the metric itself.

Example with Real Data: In our experiments we construct a global AS graph using all data from the Route Views, RIS, and BGP-mon project on June 1st 2016. This corresponds to an AS graph of more than 50k nodes with 326 viewpoints (we consider only full-feed BGP peers), and only 0.6% of all the AS paths on the Internet (16M paths out of the 2.5B). As collected paths all converge to the 326 viewpoints, ASes accommodating viewpoints and their neighboring ASes are seemingly more central than other ASes. To measure the bias obtained with real BGP data we conduct the following experiment. First, we compute the BC for all ASes from all 326 viewpoints, then we compare this distribution of BC values to BC values obtained with a smaller set of randomly selected viewpoints. The distance between two distributions is measured with...
the Kullback-Leibler divergence. Figure 2a shows that changing the number of viewpoints invariably reshapes the BC distribution, meaning that the obtained BC values are conditioned by the number of viewpoints. From these results, we hypothesize that having more than 326 viewpoints would yield different BC values thus the BC values obtained with the 326 viewpoints might not be representative of AS centrality.

3 AS HEGEMONY

We propose a new centrality metric inspired by BC but taking into account the bias of BGP viewpoints. This new metric, hereafter referred to as AS hegemony, measures the fraction of paths passing through a node for a set of selected unbiased viewpoints. To compute the hegemony of an AS $v$, we adaptively discard viewpoints with the highest, or lowest, number of paths passing through $v$. Remaining viewpoints are considered as unbiased towards $v$. Finally, the hegemony is obtained by averaging the fraction of paths going through $v$ for each remaining viewpoint.

In the following let $n$ be the total number of viewpoints, $\lfloor . \rfloor$ be the floor function and $2\alpha$ be the ratio of disregarded viewpoints. Then the AS hegemony is formally defined as:

$$H(v, \alpha) = \frac{1}{n - (2\alpha n)} \sum_{j=\lfloor \alpha n \rfloor + 1}^{n-\lfloor \alpha n \rfloor} BC_{(j)}(v)$$

(2)

where $BC_{(j)}(v)$ is the BC value computed with paths from only one viewpoint $j$ (i.e. $BC_{(j)}(v) = 1/S \sum_{w \in V} \sigma_{jw}(v)$) and these values are arranged in ascending order such that $BC_{(1)}(v) \leq BC_{(2)}(v) \leq \cdots \leq BC_{(n)}(v)$. Figure 1 depicts the AS hegemony obtained for the simple graph with three viewpoints ($\alpha = .34$). Unlike the sampled BC, the AS hegemony is consistent for each type of node: transit (.58), regional ISP (.25) and stub AS (.08).

Preliminary Findings: As we did previously with BC, we compute from real BGP data the AS hegemony using all viewpoints then we compare these results to those obtained with a lower number of randomly selected viewpoints. Figure 2a shows that the hegemony values with 20 or more viewpoints are very similar to the ones obtained from all the peers, hence the AS hegemony is more robust than BC to sampling. Note that we randomly select peers across different projects (e.g. Route Views, RIS, BGPMonitor) to obtain a diverse set of viewpoints. We found that selecting viewpoints from the same BGP collector usually yield poor results. For instance, the collector route-view4 (rv4 in Fig.2b) consists of 33 viewpoints but results obtained only with this collector are greatly diverging from the ones obtained with the 326 viewpoints. Using 20 randomly selected viewpoints provides much better results. For other collectors, route-view2 and LINX collectors provide the most diverse peers.

Further Directions: For the above results we computed the entire AS graph, but the AS hegemony is also suitable to monitor smaller graphs. For example an operator is primarily interested in finding anomalies in routes towards its own network. To focus on a certain part of the Internet we construct a smaller AS graphs using only paths bound to a set of prefixes. Such graph represents all the routes from the viewpoints to these prefixes, and high hegemony values stand for the main ASes crossed to access these prefixes. Figure 2c illustrates the hegemony distribution for different AS graphs constructed with paths to prefixes mapped to distinct countries. Data points on the right hand side of the figure depict most central ASes for these countries. For Cuba, North Korea and Pakistan we observe a few ASes with an hegemony close to 1 meaning that all paths to these countries cross central ASes. The U.S. appeared to be the country where hegemony values are the most balanced. We found that the distribution of hegemony values is usually stable over time, and significant changes are a good indication of fundamental route changes usually attributed to BGP leaks or hijacks. Furthermore, the precision and robustness of AS hegemony enable us to accurately monitor very local changes in the AS graph.

REFERENCES