Quantifying Community Evolution in Developer Social Networks: Proof of Indices' Properties

LIANG WANG, State Key Laboratory for Novel Software Technology, Nanjing University, China
YING LI, State Key Laboratory for Novel Software Technology, Nanjing University, China
JIERUI ZHANG, State Key Laboratory for Novel Software Technology, Nanjing University, China
XIANPING TAO, State Key Laboratory for Novel Software Technology, Nanjing University, China

The document provides the proof to properties of community evolution indices including community *split* and *shrink* in paper: Liang Wang, Ying Li, Jierui Zhang, and Xianping Tao. 2022. QuantifyingCommunity Evolution in Developer Social Networks. InProceedings of the30th ACM Joint European Software Engineering Conference and Symposiumon the Foundations of Software Engineering (ESEC/FSE '22), November 14–18, 2022, Singapore, Singapore.ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3540250.3549106. Proof to properties of community *merge* and *expand* is similar.

1 BRIEF INTRODUCTION TO THE PROPERTIES OF COMMUNITY SPLIT AND SHRINK INDICES

Let $I_{c_{t,i}}^{\psi}$ and $I_{c_{t,i}}^{\eta}$ denote the community *split* and *shrink* indices, respectively. Without loss of generality, we assume $m \geq 1$. The properties of the two indices are as follows.

P-1. $I_{c_{t,i}}^{\psi}$ and $I_{c_{t,i}}^{\eta}$ are <u>strictly monotonic increasing</u> functions of m, given $0 < \eta_i < 1$, and $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$.

P-2. $I_{c_{t,i}}^{\psi}$ / $I_{c_{t,i}}^{\eta}$ is a <u>strictly monotonic decreasing / increasing</u> function of η_i, respectively, for $\eta_i > 0$, given $m > 1$, and member migration distribution $\hat{\psi}_{i,j}$, $j = 1, 2, \cdots, m$ with $\mathcal{H}_{c_{t,i}} > 0$.

P-3. Given m and η_i, the <u>maximum split index</u> $I_{c_{t,i}}^{\psi} = (1 - \eta_i)\mathcal{H}_{t \to t+1}^{*}$ is obtained when the members of $c_{t,i}$ migrate to the communities detected in the next step with a even distribution, i.e., when we have $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$. And the <u>minimum split index</u> $I_{c_{t,i}}^{\psi} = 0$ is obtained when $m = 1$ or all the members of $c_{t,i}$ who stay in the project migrate to a single community in the next step, i.e., there exists a j'-th community in time $t + 1$ that $\hat{\psi}_{i,j'} = 1$ and $\hat{\psi}_{i,j} = 0, \forall j \neq j'$, resulting in $\mathcal{H}_{c_{t,i}} = 0$ and $I_{c_{t,i}}^{\psi} = 0$.

P-4. Given $m > 1$ and η_i, the <u>maximum shrink index</u> $I_{c_{t,i}}^{\eta} = \eta_i \mathcal{H}_{t \to t+1}^{*}$ is obtained when the corresponding split index is minimized, i.e., all stayed members of community $c_{t,i}$ migrate to a single community in the next step. And the <u>minimum shrink index</u> $I_{c_{t,i}}^{\eta} = \eta_i^2 \mathcal{H}_{t \to t+1}^{*}$ is obtained when the members of $c_{t,i}$ migrate evenly to communities in time $t + 1$. For the special case of $m = 1$, the shrink index is only determined by η_i.

Authors' addresses: Liang Wang, wl@nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, 163 Xianlin Ave., Nanjing, China; Ying Li, mg21070006@smail.nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, 163 Xianlin Ave., Nanjing, China; Jierui Zhang, jieruizhang@smail.nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, 163 Xianlin Ave., Nanjing, China; Xianping Tao, txp@nju.edu.cn, State Key Laboratory for Novel Software Technology, Nanjing University, 163 Xianlin Ave., Nanjing, China.
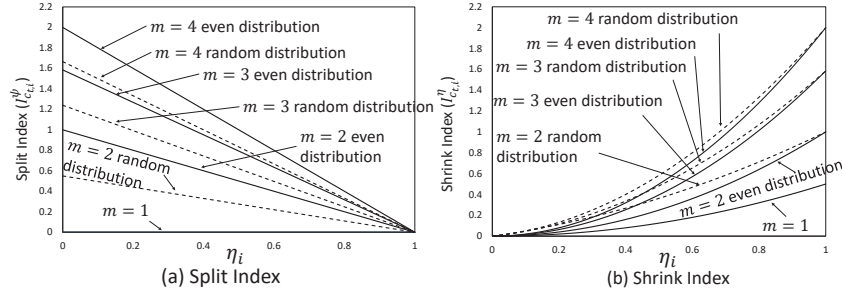
Fig. 1. Curves of the *split* and *shrink* indices under different conditions specified by m, η_i, and $\hat{\psi}_{i,j}$. The even distribution corresponds to case $\hat{\psi}_{i,j} = \frac{1}{m}, j = 1, \cdots, m$. The random distribution is obtained by randomly assigning $\hat{\psi}_{i,j}$'s values.

Fig. 1 illustrates the curves of the split and shrink indices under different conditions, from which we can find correspondence to the above properties.

2 PROOF OF THE PROPERTIES

The proof of the above four properties are as follows.

P-1. $\mathcal{I}_{c_{t,i}}^{\psi}$ and $\mathcal{I}_{c_{t,i}}^{\eta}$ are <u>strictly monotonic increasing</u> functions of m, given $0 < \eta_i < 1$, and $\hat{\psi}_{i,j} = \frac{1}{m}, j = 1, 2, \cdots, m$.

Proof. First, for community split index $\mathcal{I}_{c_{t,i}}^{\psi}$, take the partial derivative of $\mathcal{I}_{c_{t,i}}^{\psi}$ with respect to m gives us:

$$\frac{\partial \mathcal{I}_{c_{t,i}}^{\psi}}{\partial m} = \frac{\partial (1 - \eta_i) \mathcal{H}_{c_{t,i}}}{\partial m} = (1 - \eta_i) \frac{\partial \mathcal{H}_{c_{t,i}}}{\partial m}. \tag{1}$$

When $\hat{\psi}_{i,j} = \frac{1}{m}, j = 1, 2, \cdots, m$, we have $\mathcal{H}_{c_{t,i}} = -\sum_{j=1}^{m} \hat{\psi}_{i,j} \log_2(\hat{\psi}_{i,j}) = -\log_2 \frac{1}{m}$. As a result, we have

$$\frac{\partial \mathcal{H}_{c_{t,i}}}{\partial m} = \frac{1}{m \ln(2)}. \tag{2}$$

Substitute Eq. (2) into Eq. (1), we have

$$\frac{\partial \mathcal{I}_{c_{t,i}}^{\psi}}{\partial m} = \frac{1 - \eta_i}{m \ln(2)}. \tag{3}$$

We can see from Eq. (3) that $\frac{\partial \mathcal{I}_{c_{t,i}}^{\psi}}{\partial m} > 0$, for $m \geq 1$, and $0 \leq \eta_i < 1$.

As a result, the community split index $\mathcal{I}_{c_{t,i}}^{\psi}$ is a <u>strictly monotonic increasing</u> function of m, for $m \geq 1$, $0 \leq \eta_i < 1$, and $\hat{\psi}_{i,j} = \frac{1}{m}, j = 1, 2, \cdots, m$.

For the case of $\eta_i = 1$, we always have $\mathcal{I}_{c_{t,i}}^{\psi} = 0, \forall m \geq 1$, and regardless the distribution of $\hat{\psi}_{i,j}, j = 1, 2, \cdots, m$, which is intuitively correct that a community cannot split if all of its members leaves the project.

Next, for community shrink index $\mathcal{I}_{c_{t,i}}^{\eta}$, we also take its partial derivative with respect to m, which is

$$\frac{\partial \mathcal{I}_{c_{t,i}}^{\eta}}{\partial m} = \frac{\partial \eta_i (\mathcal{H}_{t \rightarrow t+1}^{*} - \mathcal{I}_{c_{t,i}}^{\psi} + \sigma_{\eta_i})}{\partial m} = \eta_i \left(\frac{\partial \mathcal{H}_{t \rightarrow t+1}^{*}}{\partial m} - \frac{\partial \mathcal{I}_{c_{t,i}}^{\psi}}{\partial m} \right). \tag{4}$$

Since $\frac{\partial \mathcal{H}_{t\to t+1}^*}{\partial m} = \frac{\partial -\log_2\left(\frac{1}{m}\right)}{\partial m} = \frac{1}{m\ln(2)}$, and $\frac{\partial \mathcal{I}_{c_{t,i}}^\psi}{\partial m} = \frac{1-\eta_i}{m\ln(2)}$ following Eq. (3), we have

$$\frac{\partial \mathcal{I}_{c_{t,i}}^\eta}{\partial m} = \frac{\eta_i^2}{m\ln(2)}. \tag{5}$$

From Eq. (5), we have $\frac{\partial \mathcal{I}_{c_{t,i}}^\eta}{\partial m} > 0$ for $m \geq 1$, $0 < \eta_i \leq 1$, and $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$.

For the case of $\eta_i = 0$, we always have $\mathcal{I}_{c_{t,i}}^\eta = 0, \forall m \geq 1$, given any member migration distribution $\hat{\psi}_{i,j}$, $j = 1, 2, \cdots, m$, which is reasonable because a community does not shrink if none of its members leave the project, regardless the number of communities detected in the next step (i.e., m).

Combining the above two results, we show that: $\mathcal{I}_{c_{t,i}}^\psi$ and $\mathcal{I}_{c_{t,i}}^\eta$ are <u>strictly monotonic increasing</u> functions of m, given $0 < \eta_i < 1$, and $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$.

□

P-2. $\mathcal{I}_{c_{t,i}}^\psi$ / $\mathcal{I}_{c_{t,i}}^\eta$ is a <u>strictly monotonic decreasing</u> / <u>increasing</u> function of η_i, respectively, for $\eta_i > 0$, given $m > 1$, and member migration distribution $\hat{\psi}_{i,j}$, $j = 1, 2, \cdots, m$ with $\mathcal{H}_{c_{t,i}} > 0$.

Proof. First, we show that community split index $\mathcal{I}_{c_{t,i}}^\psi$ is a <u>strictly monotonic decreasing</u> function of η_i under the given conditions. The partial derivative of $\mathcal{I}_{c_{t,i}}^\psi$ with respect to η_i is

$$\frac{\partial \mathcal{I}_{c_{t,i}}^\psi}{\partial \eta_i} = \frac{\partial (1 - \eta_i)\mathcal{H}_{c_{t,i}}}{\partial \eta_i} = -\mathcal{H}_{c_{t,i}}. \tag{6}$$

As a result, we have $\frac{\partial \mathcal{I}_{c_{t,i}}^\psi}{\partial \eta_i} < 0$, meaning that $\mathcal{I}_{c_{t,i}}^\psi$ is a <u>strictly monotonic decreasing</u> function of η_i, as long as $\mathcal{H}_{c_{t,i}} > 0$.

We have $\mathcal{H}_{c_{t,i}} = 0$ and thus $\mathcal{I}_{c_{t,i}}^\psi = 0, \forall \eta_i \in [0, 1]$ when only one of the $\hat{\psi}_{i,j}$'s is one with rest of the $\hat{\psi}_{i,j}$'s equal to zero (including the case that $m = 1$). Intuitively, this case means that a community is not regarded as splitting if all of its remaining members migrate to a single community in the next step, regardless the amount of members who leave the project.

Next, we show that community shrink index $\mathcal{I}_{c_{t,i}}^\eta$ is a <u>strictly monotonic increasing</u> function of η_i under the given conditions. Taking the partial derivative of $\mathcal{I}_{c_{t,i}}^\eta$ with respect to η_i gives us

$$\frac{\partial \mathcal{I}_{c_{t,i}}^\eta}{\partial \eta_i} = \frac{\partial \eta_i (\mathcal{H}_{t\to t+1}^* - \mathcal{I}_{c_{t,i}}^\psi + \sigma_{\eta_i})}{\partial \eta_i} = (\mathcal{H}_{t\to t+1}^* - \mathcal{I}_{c_{t,i}}^\psi + \sigma_{\eta_i}) + \eta_i \frac{\partial (\mathcal{H}_{t\to t+1}^* - \mathcal{I}_{c_{t,i}}^\psi + \sigma_{\eta_i})}{\partial \eta_i}$$
$$= \mathcal{H}_{t\to t+1}^* - (1 - \eta_i)\mathcal{H}_{c_{t,i}} + \sigma_{\eta_i} - \eta_i \frac{\partial \mathcal{I}_{c_{t,i}}^\psi}{\partial \eta_i} + \eta_i \frac{\partial \sigma_{\eta_i}}{\partial \eta_i}. \tag{7}$$

The forth term in Eq. (7) is $\eta_i \frac{\partial \mathcal{I}_{c_{t,i}}^\psi}{\partial \eta_i} = \eta_i \frac{\partial (1-\eta_i)\mathcal{H}_{c_{t,i}}}{\partial \eta_i} = -\eta_i \mathcal{H}_{c_{t,i}}$. And the last term in Eq. (7) is $\eta_i \frac{\partial \sigma_{\eta_i}}{\partial \eta_i} = 0.5\eta_i$ when $m = 1$, and $\eta_i \frac{\partial \sigma_{\eta_i}}{\partial \eta_i} = 0$ when $m > 1$ following the definition of σ_{η_i}.

For the case of $m > 1$, Eq. (7) can be written as

$$\frac{\partial \mathcal{I}_{c_{t,i}}^{\eta}}{\partial \eta_i} = \mathcal{H}_{t \to t+1}^* - (1 - 2\eta_i)\mathcal{H}_{c_{t,i}}. \tag{8}$$

Because $\mathcal{H}_{t \to t+1}^*$ is the maximum entropy, we have $\mathcal{H}_{t \to t+1}^* > 0$, $\mathcal{H}_{t \to t+1}^* \geq \mathcal{H}_{c_{t,i}} \geq 0$, and $(1 - 2\eta_i) < 1$ when $\eta_i > 0$. We then have $\frac{\partial \mathcal{I}_{c_{t,i}}^{\eta}}{\partial \eta_i} > 0$ for $\eta_i > 0$ and $m > 1$.

For the case of $m = 1$, we have $\sigma_{\eta_i} = \eta_i \frac{\partial \sigma_{\eta_i}}{\partial \eta_i} = 0.5\eta_i$, and $\mathcal{H}_{t \to t+1}^* = \mathcal{H}_{c_{t,i}} = 0$. Eq. (7) then becomes

$$\frac{\partial \mathcal{I}_{c_{t,i}}^{\eta}}{\partial \eta_i} = \eta_i. \tag{9}$$

We also have $\frac{\partial \mathcal{I}_{c_{t,i}}^{\eta}}{\partial \eta_i} > 0$ for $\eta_i > 0$ and $m = 1$. As a result, $\mathcal{I}_{c_{t,i}}^{\eta}$ is a <u>strictly monotonic increasing</u> function of η_i for $\eta_i > 0$, and $m \geq 1$.

Summarizing the above, we show that: $\mathcal{I}_{c_{t,i}}^{\psi} / \mathcal{I}_{c_{t,i}}^{\eta}$ is a <u>strictly monotonic decreasing / increasing</u> function of η_i, respectively, for $\eta_i > 0$, given $m > 1$, and the distribution of member migration $\hat{\psi}_{i,j}$, $j = 1, 2, \cdots, m$ with $\mathcal{H}_{c_{t,i}} > 0$. □

P-3. Given m and η_i, the <u>maximum split index</u> $\mathcal{I}_{c_{t,i}}^{\psi} = (1 - \eta_i)\mathcal{H}_{t \to t+1}^*$ is obtained when the members of $c_{t,i}$ migrate to the communities detected in the next step with a even distribution, i.e., when we have $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$. And the <u>minimum split index</u> $\mathcal{I}_{c_{t,i}}^{\psi} = 0$ is obtained when $m = 1$ or all the members of $c_{t,i}$ who stay in the project migrate to a single community in the next step, i.e., there exists a j'-th community in time $t + 1$ that $\hat{\psi}_{i,j'} = 1$ and $\hat{\psi}_{i,j} = 0, \forall j \neq j'$, resulting in $\mathcal{H}_{c_{t,i}} = 0$ and $\mathcal{I}_{c_{t,i}}^{\psi} = 0$.

Proof. $\mathcal{I}_{c_{t,i}}^{\psi} = (1 - \eta_i)\mathcal{H}_{c_{t,i}}$ is a monotonic increasing function of $\mathcal{H}_{c_{t,i}}$ for $0 \leq \eta_i \leq 1$. Referring to the properties of information entropy [1], entropy $\mathcal{H}_{c_{t,i}} = \mathcal{H}_{t \to t+1}^* = -\log_2 \frac{1}{m}$ is a maximum when $\hat{\psi}_{i,j} = \frac{1}{m}$, $j = 1, 2, \cdots, m$. And the minimum value of $\mathcal{H}_{c_{t,i}} = 0$ is obtained when only one of the $\hat{\psi}_{i,j}$'s equals to one and others equal to zero, which also includes the case of $m = 1$. As a result, the maximum split index is $\mathcal{I}_{c_{t,i}}^{\psi} = (1 - \eta_i)\mathcal{H}_{t \to t+1}^* = -(1 - \eta_i)\log_2 \frac{1}{m}$ given η_i. When $\eta_i = 0$, we have $\mathcal{I}_{c_{t,i}}^{\psi} = -\log_2 \frac{1}{m}$ being the maximum possible value for the split index, which is only determined by m. And the minimum split index is $\mathcal{I}_{c_{t,i}}^{\psi} = 0$. □

P-4. Given $m > 1$ and η_i, the <u>maximum shrink index</u> $\mathcal{I}_{c_{t,i}}^{\eta} = \eta_i \mathcal{H}_{t \to t+1}^*$ is obtained when the corresponding split index is minimized, i.e., all stayed members of community $c_{t,i}$ migrate to a single community in the next step. And the <u>minimum shrink index</u> $\mathcal{I}_{c_{t,i}}^{\eta} = \eta_i^2 \mathcal{H}_{t \to t+1}^*$ is obtained when the members of $c_{t,i}$ migrate evenly to communities in time $t + 1$. For the special case of $m = 1$, the shrink index is only determined by η_i.

Proof. Given $m > 1$ and η_i, the community shrink index $\mathcal{I}_{c_{t,i}}^{\eta} = \eta_i(\mathcal{H}_{t \to t+1}^* - \mathcal{I}_{c_{t,i}}^{\psi} + \sigma_{\eta_i})$ is a monotonic decreasing function of the community split index $\mathcal{I}_{c_{t,i}}^{\psi}$. Referring to Property 3 presented above, we have the shrink index maximized when the split index is minimized, i.e., $\mathcal{H}_{c_{t,i}} = 0$. The maximum shrink index is given by $\mathcal{I}_{c_{t,i}}^{\eta} =$

$\eta_i \mathcal{H}^*_{t \to t+1} = -\eta_i \log_2 \frac{1}{m}$. And the minimum shrink index $\mathcal{I}^\eta_{c_{t,i}} = \eta_i^2 \mathcal{H}^*_{t \to t+1} = -\eta_i^2 \log_2 \frac{1}{m}$ when $\mathcal{I}^\psi_{c_{t,i}} = (1 - \eta_i)\mathcal{H}^*_{t \to t+1}$.

In the above analysis we have $\sigma_{\eta_i} = 0$ because we assume $m > 1$. When $m = 1$, we have $\sigma_{\eta_i} = 0.5\eta_i$, and $\mathcal{H}^*_{t \to t+1} = \mathcal{H}_{c_{t,i}} = 0$. The shrink index is $\mathcal{I}^\eta_{c_{t,i}} = 0.5\eta_i^2$, which is only determined by η_i.

Following the above analysis, if we consider the change of η_i, we can find that the maximum and minimum possible value of the shrink index is $\mathcal{I}^\eta_{c_{t,i}} = -\log_2 \frac{1}{m}$ when $\eta_i = 1$, and $\mathcal{I}^\eta_{c_{t,i}} = 0$ when $\eta_i = 0$, respectively.

$\square$

From properties **P-3** and **P-4** we can further see that given $m > 1$, the community split and shrink indices vary in the same range given by $\mathcal{I}^\psi_{c_{t,i}} \in [0, -\log_2 \frac{1}{m}]$, and $\mathcal{I}^\eta_{c_{t,i}} \in [0, -\log_2 \frac{1}{m}]$, with different values of η_i and the distribution of member migration specified by $\hat{\psi}_{i,j}, j = 1, 2, \cdots, m$. As a result, it is feasible for us to draw meaningful results, such as *the community shows a stronger trend of splitting / shrinking*, by directly comparing the values of community split and shrink indices.

ACKNOWLEDGMENTS

REFERENCES

[1] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.