Identifying research facilitators in an emerging Asian Research Area

**Philip S. Cho · Huy Hoang Nhat Do · Muthu Kumar Chandrasekaran ·
Min-Yen Kan**

Abstract   We introduce a novel set of metrics for triadic closure among individuals or groups to model how co-authorship networks become more integrated over time. We call this process of triadic, third-party mediated integration, *research facilitation*. We apply our research facilitation or RF-metrics to the development of the Pan-Asian SNP (PASNP) Consortium, the first inter-Asian genomics network. Our aim was to examine if the consortium catalyzed research facilitation or integration among the members and the wider region. The PASNP Consortium is an ideal case study of an emerging Asian Research Area because its members themselves asserted a regional Asian identity. To validate our model, we developed data mining software to extract and match full author and institutional information from the PDFs of scientific papers.

What is Asia in scientometrics research?

Despite growing interest about Asia as a research area, scientometrics as a field has yet to critically address how to define and conceptualize the region. A survey of recent literature reveals no agreement over what countries represent or belong to Asia and what historical,

P. S. Cho (✉)
Asia Research Institute, National University of Singapore, Singapore
e-mail: aripcss@nus.edu.sg

H. H. N. Do · M. K. Chandrasekaran · M.-Y. Kan
Department of Computer Science, National University of Singapore, Singapore
e-mail: kanmy@comp.nus.edu.sg

M.-Y. Kan
Interactive and Digital Media Institute, National University of Singapore, Singapore

🖄 Springer

cultural, or economic forces are driving inter-Asian scientific collaboration. For example, in Haustein et al.'s (2011) study of an Asia–Pacific research area, "the selection was purely subjective" (p. 729). This otherwise thoughtful study represents Asia through a range of culturally, economically, and geographically diverse countries including China, New Zealand and Malaysia, but notably excludes others such as India or Mongolia. In contrast, researchers such as Kumar et al. (2011) and Basu (2001) stress that India is essential to understanding the rise of Asia. Garg (2002) and Guan and Ma (2007) choose India and China as the core of the region. Additional studies conceptualize the imaginary boundaries of Asia as extremely broad, encompassing nearly a third of the world. Khan and Park's (2012) review of triple helix in Asia, confusingly includes Singapore (located south of India and west of the Philippines) among *East Asian* countries (p. 2). Whereas, Shin et al.'s (2012) analysis of triple helix, bounds Asia with East Asian countries like Japan and China on one side, and Gulf countries like Saudi Arabia, "on the other side of Asia" (p. 311).

Scientometrics scholars have thus been using the concept of Asia as a variable label about the identity of the region. How scientometrics publications represent Asia has paid little attention to how the governments, institutions, or individuals under study might see and group themselves. We feel that the concept of Asia should be treated critically as a flexible social construction rather than a natural kind.

In this paper we:

- Critique the history of how scientometrics research has described and represented Asia, largely through symmetric and asymmetric dyads of nations.
- Introduce a novel set of metrics for triadic closure among individuals or groups to model how co-authorship social networks become more integrated over time. We call this process of triadic, third-party mediated integration as *research facilitation* and rigorously derive mathematical definitions of the variety of research facilitation.
- Develop data mining software to gather, extract, match, and disambiguate the names of all authors and institutions from the PDFs of scientific papers.
- Apply our RF-metrics and data mining software to analyze the evolution of the co-authorship network of the PASNP Consortium from 2004 to 2011 (The HUGO Pan-Asian SNP Consortium 2009). The PASNP, the first inter-Asian genomics consortium consisted of 92 core researchers at 40 institutions in 11 self-professed Asian countries. The broader network included 4,261 researchers in 45 countries. The consortium's major finding, "Mapping Human Genetic Diversity in Asia," was published in *Science*, 2009. Using data on single nucleotide polymorphisms (SNPs), the paper showed how all Asians shared a common genetic ancestry and populated the region through a single wave of northward migration (Cho et al. 2013; Do et al. 2013). We chose the PASNP as a case study in the emergence of an Asian Research Area, because its participants defined themselves as, "an Asian scientific network," (Murthy 2009) that was, "conceived by Asians in Asia and executed, funded, and completed by an Asian consortium" (Normile 2004). In other words, the PASNP Consortium members *reflexively* asserted an Asian identity, rather than we, the authors of this paper, imposing one on them.

An historical critique of scientometric representations of Asia as dyadic co-authorship networks

Scientometric analyses of co-authorship in Asia have primarily focused on dyadic, paired, relationships between two countries to construct different narratives of a world scientific

center and periphery (Shils 1975; Schott 1988). These representations of Asia have conflicted depending on researchers' narrative aims in calculating the collaborative strength between two countries as symmetric or asymmetric.

Symmetric measures such as Salton's cosine favor countries having high publication rates (Salton and McGill 1983). Even now, this often inherently makes ties to advanced Western countries the most significant. Shils (1975) and later Schott (1988) asserted an uneven distribution of scientific capital between a North American and Western European core of scientific activity, with Asia and the rest of the world at the periphery.

As the strongest ties were to a Western center, these studies rarely found Asia as an integrated research area for intra-regional scientific collaboration. Applying Salton's measure on Science Citation Index (SCI) data from 1981 to 1985 for 167 countries, Schubert and Braun (1990) found no significant "intrinsic cooperativity" among Asian countries such as Japan and China. In a related work, Okubo et al. (1992) used a correspondence factorial analysis (CFA) between country pairs and a minimum spanning tree to analyze 98 countries from 1981 to 1986. They found that Asian countries had different patterns of scientific collaboration. Japan was isolated from Asian–Pacific countries and more closely tied to the US. Korea lay at the periphery, while China and India were mostly low scientific producers.

Asymmetric measures of the relative collaborative strength between two countries aimed to better reflect the value-laden choice of regional partners. According to Frame and Carpenter (1979), science was "not value free and transcending national boundaries … [but] tied to nationality, geographical locale, linguistic, cultural, and political [factors]." Examining the same period as Okubo et al. (1992) for SCI data on 97 countries, Luukkonen et al. (1992) mapped a different pattern of international scientific co-authorship emphasizing inter-Asian collaboration. They criticized Schubert and Braun's use of Salton's measure as underestimating the strength of links between small, science-producing countries. To address this problem, they employed a Weighted Affinity Index (WAI) (Leclerc and Gagné 1994) to account for asymmetric relationships in the relative importance of collaboration between two countries. While the study found that the United States was the most important co-authorship partner for Japan in terms of absolute number, in relative terms, Japan was closely tied to its Asian neighbors of China and India. Using the WAI, Leclerc and Gagné (1994) drew a map of 1990 SCI data representing a Nippon Asiatic Zone in a global triad of international scientific cooperation. Along with the US and Western Europe, Japan became a new core with the rest of Asia as the periphery. Recent studies, as we reviewed in the introduction, continue to use dyadic measures to construct narratives of Asia within a multipolar world of science.

Toward representations of Asia using triadic closure in co-authorship networks

We argue that a more accurate depiction of Asia needs to use triadic measures to better model how co-authorship social networks in the region are becoming more integrated. Our aim is to identify the key players in facilitating collaborative research in an emerging pan-Asian research network. To achieve this goal, we developed metrics to measure the impact of research facilitators on scholarly publication networks. Informally, we measure the degree of *facilitation* demonstrated by an individual researcher by assessing how often they complete the pathway or connection between two other researchers who subsequently co-author together. Rather than dyadic relationships, this role of the facilitator leads to a

triadic closure over time. This is when the pathway between three actors is closed in a triangle relationship.

Theorists in international relations, economics, and innovation studies have increasingly turned to triadic measures in network analysis of regional interdependence. They note that measuring only direct collaborative ties between two countries is inadequate to understand their interdependence in forming an integrated region. Specifically, such pairwise analyses of dyads are ideal for core-periphery or hub models. However, the interplay that defines entire regions necessitates analytical methods that account for not just direct, but also indirect, mediated, and facilitated interactions in the formation of cliques, whose actors interact among each other.

Maoz (2009) notes that focusing on only direct trade ties ignores and oversimplifies the costs of regional economic interdependence; extra-dyadic relationships deter conflict even in cases where the countries have little direct trade. Burt (2000), Obstfeld (2005) and Wineman et al. (2009) have pointed to third-party knowledge brokers who connect two groups and create opportunities for innovative combinations of expertise and information by closing structural holes between them.

Mathematical definitions of research facilitation as triadic closure among individuals or groups

We propose a set of research facilitation or RF-metrics to identify researchers who play critical roles over time in facilitating collaboration among other individuals, groups, or countries. We seek to discover actors who complete a social networking pathway between two researchers or groups before these two collaborate. This involves a temporally sequenced triadic relationship in which a researcher C has co-authored with two others, A and B, at some time before A and B begin collaborating (Fig. 1). Research facilitators, thus, mediate greater interconnectivity within a network, rather than simply act as a central hub. This approach departs from traditional dyadic co-authorship metrics that identify and rank researchers based on their total connections or paths within static networks.

Previously, Rapoport (1953) and Granovetter (1973) had theorized that two actors' tie strength varies directly with the number of their shared acquaintances and friends. Kossinets and Watts (2006) empirically showed how the probability of triadic closure is influenced by social contexts such as shared classes. We introduce a series of metrics to formally describe triadic closure in scientific collaboration and to rank key researchers as facilitators.

We formally define a set of properties that build up to our RF-metric for identifying research facilitators using triadic closure. These RF-metrics can analyze mediated interconnectivity and integration at different levels of granularity including individuals, institutions, groups, and countries.

Definitions 1 and 2 formalize the primitives of a scholarly collaboration network over a period of time.

Definition 1 A time unit co-authorship graph $G_t = (V, E)$ is a graph with a set of vertices V (or equivalently V nodes) that represents individual researchers and a set of weighted edges $E = (v_1, v_2)$ that represents co-authoring relations between the researchers that occur during a time unit t.
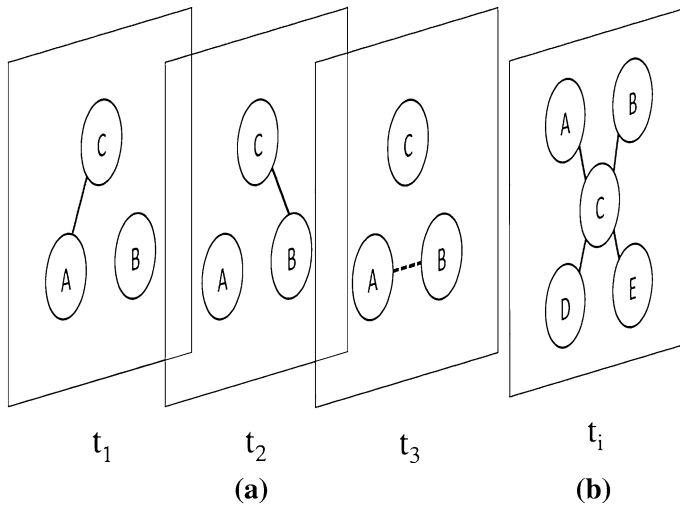
Fig. 1 Facilitators and hubs. (**a**) In order for C to be a facilitator for A and B, AB must collaborate after CA and CB; (**b**) C is the hub of A, B, D, and E, none of which have necessarily collaborated. We use a dashed line to indicate the publication that is facilitated.

Definition 2 A co-authorship graph time series G_T consists of an ordered set of G_t where $(i < t \leq j)$. We note that the vertices in each of the graphs G_i are conserved between time intervals; only edges change (indicating co-authorship during that particular time unit i).

In this graph formalism, researchers are nodes, and thus a publication with k researchers would manifest as a complete k subcomponent of the graph. Singly-authored publications (e.g., monographs) are not represented in the graph. Edges are assigned weights according to how many publications the two researchers co-authored during t. In our analysis, we also take t to be measured in years.

Definition 3 The first-time collaboration $\overleftarrow{(v_1, v_2)}(t)$ between two researchers v_1, v_2 is defined as the earliest edge $e \in G_t$ that exhibits a co-author relation between v_1 and v_2.

For notational convenience to emphasize the researcher-centric nature of facilitation, we also use capital letters to denote individual researchers. We use $\overleftarrow{AB}(t)$ to be equivalent notation to $\overleftarrow{(v_1, v_2)}(t)$, where $v_1 = A$ and $v_2 = B$. Since co-authorship is symmetric, $\overleftarrow{AB}(t) = \overleftarrow{BA}(t)$.

Definition 4 A temporal path $ABC(t_1, t_2)$ is equivalent to $e_1 = (A, B) \in G_{t_1}$ and $e_2 = (B, C) \in G_{t_2}$.

Our notation captures the possibility that a complete co-authorship path may form the path over different time units. In previous works on evolving networks, such as Tang et al. (2012), paths exist as a whole over time, rather than constructed individually piecewise, forming a whole over time.

Given these primitives, we can define a facilitator as an individual who may have introduced two researchers to each other, which results in a co-authored publication. Facilitation is a triadic measure in which the co-authorship AB closes a loop among all three individuals. We can now define *facilitators* and *prime facilitators*:

Definition 5 C is a *facilitator* of $\overleftarrow{AB}(t)$ iff there exists $\overleftarrow{CA}(t_1)$ and $\overleftarrow{CB}(t_2)$ such that $max(t_1, t_2) \leq t$. We write *facil(C, AB)*, or equivalently *facil(C, BA)*, to denote that C facilitates AB.

Definition 6 C is a *prime facilitator* of AB iff C is a facilitator of AB; furthermore, given the temporal path $ACB(t_1,t_2)$ for all $AC'B(t_3, t_4)$, $max(t_1, t_2) \leq max(t_3, t_4) < t$. We write *prime(C, AB)* to denote that C is the first to facilitate, or prime facilitates, AB.

Two co-authors have a facilitated relationship if they share a common co-author from some past time. Some co-authoring pairs may thus have multiple facilitators, a set of people with whom they commonly published. Other co-authoring pairs may have none, not sharing a single co-author from the past.

Informally, a prime facilitator is the earliest, common contact that two collaborators have in the past. Prime facilitation is not necessarily a unique property, as several individuals may tie as first to have published with both members of a co-authoring pair, at various times in the past. For example, if $max(t_1, t_2) = max(t_3, t_4)$, then C and C' are both prime facilitators of $\overleftarrow{AB}(t)$ as shown in Fig. 2. Note that we exclude the case when three researchers who have never co-authored with one another in the past, co-author for the first time in one year to form a triangle.
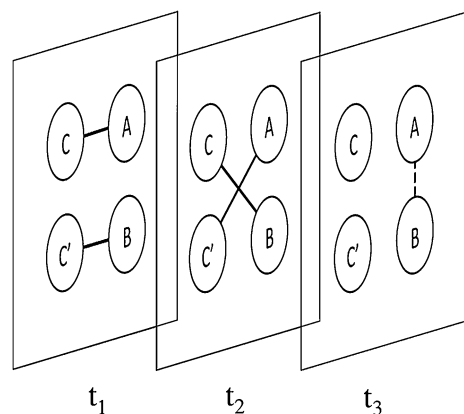
We recognise a few basic assumptions. Labeling an individual C as a facilitator means that C is a *potential* facilitator for A and B; from co-authorship records, there is no concrete evidence that (a) C did introduce A and B to each other, and that (b) this brokering caused A and B to co-author a publication.

Although we have no way of directly ascertaining whether an individual facilitates future collaborative ventures, we believe that there is a strong correlation between the occurrence of these graphical triads and actual facilitation. To create a facilitator metric to rank individuals, we must thus consider (a) how to estimate the strength of a particular facilitator relation, and (b) how to aggregate multiple facilitator relations to quantify the RF-metric.

Weighting individual acts of facilitation

We need to assign a score to each *facil(C, AB)* relation (as introduced in Definition 5) present in G_T. We choose to represent facilitation as a probability $p(facil(C, AB))$, such that all possible facilitators of AB partition a unit score among them (Eq. 1):

Fig. 2 Two facilitators C and C' of AB, who are both also prime facilitators. Dashed lines indicate publications that were facilitated.

t_1 t_2 t_3

$$p(facil(C, AB)) = \frac{|CA(t_1, t_i)| \cdot |CB(t_1, t_i)|}{\sum_{\forall C'}(|C'A(t_1, t_i)| \cdot |C'B(t_1, t_i)|)}, \quad \text{where } (t_1 \le t_i \le t_{max}) \qquad (1)$$

Here, $|CA(t_1, t_i)|$ is the magnitude of the set of co-author edges $CA(t_1, t_i)$, measuring the total number of collaborative papers between C and A. In other words, the sum of all edges CA from the first time unit t_1 until the establishment of the first-time collaboration $\overleftarrow{AB}(t_i)$.

Certain collaborative relationships carry more weight than others. Facilitating A and B who co-author frequently has greater impact than facilitating D and E who co-author rarely. We thus assign each collaboration a significance score $sig(AB) = |AB(t_i, t_{max})|$, which is the total number of publications between A and B from their earliest collaboration until the final time step present in G_T. Our basic RF-score for a particular tuple (C, AB) is Eq. (2):

$$\text{base-rf}(C, AB) = p(facil(C, AB)) \times sig(AB)$$

Basic research facilitation $\qquad (2)$

Extension of RF-metrics to account for groups

In an era of Big Science, researchers rarely work alone. Large projects, such as the PASNP Consortium, are common. We hypothesize that the formation of such groups can influence the importance of a facilitation act. Informally, if two groups $\mathcal{A}$ and $\mathcal{B}$ collaborate frequently, there is a high probability that A of group $\mathcal{A}$ will co-author with B of the other group $\mathcal{B}$. If C facilitated A and B's co-authorship, $rf(C, AB)$ needs to be accordingly dampened by the fact that that these groups regularly collaborate.

Thus, when grouping information about researchers is known, we want to modify our base-rf$(\cdot)$ scores to account for the known groupings. We account for this by first calculating *group temporal proximity*, or $gtp(A, B, t_i)$, which is the total number of publications between group $\mathcal{A}$ and $\mathcal{B}$ until time step t_i. We thus normalize the $rf(C, AB)$ score against the appropriate $gtp(A, B)$ in Eq. (3):

$$\text{group-rf}(C, AB) = \frac{\text{base-rf}(C, AB)}{gtp(A, B, t_i)}, \quad t_1 \le t_i \le t_{max}$$

Group research facilitation $\qquad (3)$

Recursive facilitation

An introduction of A to B by C, can in turn enable A or B to facilitate introductions of other individuals to each other at later time steps. The importance of the facilitation action developed thus far, base-rf(C, AB), should also partially be credited for the further facilitation it enabled.

We modify the basic RF-score to account for such indirect facilitation. Continuing our example, A may subsequently facilitate D to collaborate with B, and likewise, B could facilitate E to collaborate with A. These two later collaborations also should be credited to C's initial facilitation.

Our final, single-facilitation metric sums up these later facilitations and adds them to the base RF-score for the collaboration. Note that we only add the facilitation credit to cases where one of the initial collaborative researchers (i.e., A or B) are one of the collaborators facilitated in the later collaboration.

$$\text{rf}_\text{b}(C, AB) = \text{base-rf}(C, AB) + \sum_{\forall D} \text{rf}_\text{b}(A, BD) + \sum_{\forall E} \text{rf}_\text{b}(B, AE)$$

(4)

Basic research facilitation (recursive)

Correspondingly, for the group facilitation version of our metric, we replace base-rf(C, AB) with group-rf(C, AB) in Eq. (5):

$$\text{rf}_\text{g}(C, AB) = \text{group-rf}(C, AB) + \sum_{\forall D} \text{rf}_\text{g}(A, BD) + \sum_{\forall E} \text{rf}_\text{g}(B, AE)$$

(5)

Group research facilitation (recursive)

To compute a final facilitation score for an individual in G_T, we sum all $rf(\cdot)$, i.e., we add together the individual scores each facilitation act for which the individual can be responsible in Eq. (6):

$$\text{RF}_*(C) = \sum_{\forall A \forall B} \text{rf}_*(C, AB)$$

(6)

Research facilitation score

The case for non-recursive scores

For cases where tracing the recursive ramifications of facilitation is not desirable, we provide non-recursive versions of our RF-metric. Eq. (6) can be modified accordingly:

The non-recursive versions of the basic and group RF-scores are:

$$\text{RF}_\bar{\text{b}}(C) = \sum_{\forall A \forall B} \text{base-rf}(C, AB)$$

(7)

Basic research facilitation score (non-recursive)

$$\text{RF}_\text{g}(C) = \sum_{\forall A \forall B} \text{group-rf}(C, AB)$$

(8)

Group research facilitation score (non-recursive)

Software implementation

We employed automated and semi-automated methods to collect data and analyze the PASNP Consortium's network. We developed two significant software packages: the Google Scholar Network Analysis Package (GSNAP), and Enlil.

Our software pipeline harvests scholarly publications as PDFs, extracts and matches the complete list of authors and institutions, and analyzes the time-series co-authorship network using our $RF(\cdot)$ and other metrics. This work flow is illustrated in Fig. 3.

GSNAP is an application programming interface (API) for Google Scholar that collects publication records for authors. GSNAP is basically a lightweight layer to present a uniform XML interface to search results from Google Scholar. We note that much of the genomic literature is freely available and indexed by Google Scholar, which can substitute for commercial SCI metadata. However, Google Scholar's metadata often does not have complete information on authors and institutional affiliations.

Enlil solves this problem by taking a publication as input and extracting author names with their respective affiliations. As input, it takes in an XML representation of a
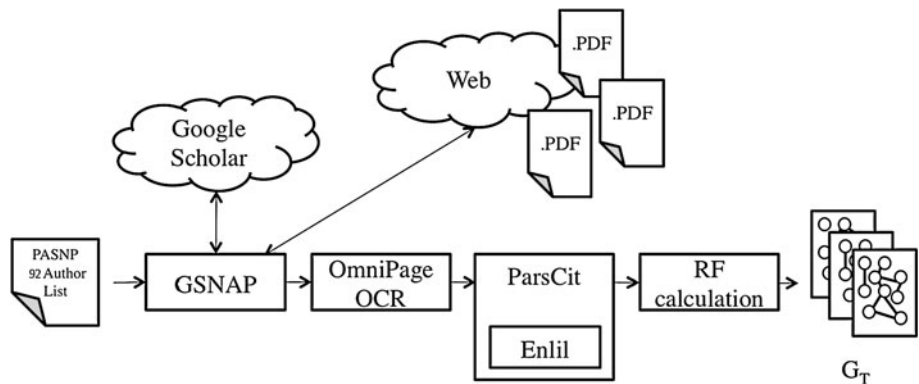
Fig. 3 Architecture of our analysis pipeline.

publication that retains the spatial placement of words and their font information (i.e., font face and size). This data is used by a pipeline of two supervised, machine-learned classifiers: a Conditional Random Field (CRF) is used to first identify and extract potential author and affiliation names, and a Support Vector Machine (SVM) is then used to match authors to their appropriate affiliations. While it may seem easy to extract and match authors with their institutions with simple heuristics, in practice there are a wide range of ways that authors and publishers express these relations (i.e., affiliations in footnotes, additional authors listed in endnotes, high-energy physics authoring teams consisting of hundreds of authors). The wide array of variations motivated our work to use supervised machine learning for Enlil.

Methodology

We now review how our software pipeline is employed to build the co-author networks G_T. From official PASNP Consortium sources, we know the names and affiliations of the 92 authors who participated in the *Science* 2009 article (The HUGO Pan-Asian SNP Consortium 2009). Since the consortium formed in 2004, we defined our time frame of analysis from 2004 to 2011. This gives us 8 one-year time steps. We then manually checked their affiliations for the time frame of the analysis, from 2004 to 2011, by visiting their web page or consulting departmental websites.

Since a large portion of the researcher names are not unique (i.e., scientists from several fields share the same namesake), merely using the researchers' names would significantly distort the resulting analysis, as authors with common names would have much larger publication counts.

We take two additional steps to filter erroneous publication records. First, we use GSNAP to query Google Scholar to obtain publication records for each of the 92 researchers. We query both researchers' names as well as their known affiliations. This works to limit the false records, eliminating those where the researcher names do not co-occur with researcher affiliations. Second, we omit specific types of Google Scholar search results that in our empirical observation generally are not considered valid scientific publications. We exclude records which do not indicate the year published, which are

published too far in the past, or which are not linked to its source journal or conference website.

For the remaining publication records, we programmatically download the Portable Document Format (.PDF) electronic copy. Publicly available papers are fetched directly from the web accessible address provided by Google Scholar. Those needing paid subscription are downloaded from the corresponding publishers through our library unit's institutional contract. To execute the latter, we manually constructed modules to download files from the majority of the publishers represented by the authors in the PASNP network: Nature, Elsevier, Wiley, PLOS, PNAS, and several others. Our goal was to ensure strong coverage of the full content links. The resulting set of 1,391 PDF documents formed the input document set used to build the PASNP Consortium co-author network.

We employed commercial optical character recognition software to render the PDF into a machine-processable XML format that preserves the spatial coordinates of the recognized words as well as font information.[1] We then ran the open-source logical document structure parsing package, ParsCit (Councill et al. 2008) with this input, modified to embed our author-affiliation extraction and matching module, Enlil (Do et al. 2013).

The final output is thus a list of the author names, and the institutional affiliation(s) for each author, for all 1,391 documents. We use these institutional affiliations to define groupings, which were used to properly compute the group-rf$(\cdot)$ component. This component is needed to compute the full $RF(\cdot)$ metric. We compute the groupings on a country basis (i.e., the country in which a researcher's institution is located). We store the information about each paper, author and country (i.e., group affiliation) into a database to facilitate the final downstream analysis.

Finally, we implemented the individual steps to calculate the component equations leading up to the to $RF(\cdot)$ metric.

Applying the RF-metric to rank individual researchers and analyze facilitation in the PASNP Consortium network

We employ the prime facilitator metric (Eq. 4) and rank individual researchers in the consortium. We focus on prime facilitators because they are the mediators of first-time collaborations. In recursive chains of facilitation over successive years, prime facilitators are on the shortest, most direct, temporal path. For small-to-medium size networks, prime facilitators are the best probe for tracing back the most direct path of facilitations over time.[2]

We do this to examine the role of facilitation in the developing co-authorship network of the PASNP Consortium. Facilitated, new collaborations can in turn lead to additional first-time collaborations. The prime facilitator metric takes the cumulative weight of these first-time interactions over time. Facilitation, thus, plays a different role in connecting a social network than betweenness centrality.

We first provide some basic statistics to generally describe the network. From 2004 to 2011, the 92 consortium members published a total of 1,391 papers with 4,261 co-authors. Table 1 provides the number of edges, author node degree, and other basic statistics on the growth of the network. The consortium's 2009 article, "Mapping Human Genetic Diversity

---

[1] Nuance Omnipage, versions 16 and 17.

[2] Otherwise, all initial members of a small networks might have some traceable path to later-facilitated connections.

in Asia," marked a pivotal year. The extended network of 4,261 researchers slightly doubled their publication rate from 114 papers in 2004 to 229 in 2009. Publications in 2009 also reflected a spike of 1,843 first-time collaborations, more than double the figure of 811 during the year before. The average number of authors per paper remained within a narrow range of 7.7 to 9.9.

Identifying and ranking prime facilitators in the PASNP network

The top 15 prime facilitators had RF-scores ranging from 399 to 3,682. Table 2 ranks the top 15 authors by their RF-score and compares it with their corresponding betweenness centrality metric. On average, authors in the entire network had an RF-score of 37. Recall that the RF-score for an individual in a network is a factor of the number of facilitations (cf. Eq. 6), the importance of each facilitation (cf. Eq. 1) and the recursive impact of those facilitations (cf. Eq. 4). Which of these factors contributed the most to the resulting RF-scores, and how does that impact our assessment of the researchers involved?

The significance of recursion on a researcher's facilitation score can be observed as the ratio of $RF_b(C)$ (cf. Eq. 7) to $RF_{\bar{b}}(C)$ (cf. Eq. 6). Recursion is less important as the ratio increases. Table 3 shows that recursion significantly adds by almost two-thirds to the average score of top research facilitators.

We also discovered that the ratio reflects the ground truth about the status of researchers. In Table 3, the two researchers with the highest ratios were Edision Liu (0.64) and Guoping Zhao (0.80). Edision Liu was the principal investigator of the overall PASNP Consortium, as well as president of the Human Genome Organization (HUGO). Guoping Zhao was principal investigator in China. We hypothesize that as principal investigators, these two researchers scored highly as facilitators because of their centrality in the network rather than their continued involvement or impact in recursive chains of collaborative work.

Other researchers such as Wei Huang or Li Jin score highly as facilitators due to both their centrality in the network and their recursive continued impact. Wei Huang led a tightly-knit group including Ying Wang and Wentao Yuan. They worked closely with another team led by Li Jin to collect and analyze genetic samples. We hypothesize that researchers such as Wei Huang and Li Jin score highly as facilitators because they lead small groups whose members frequently collaborate, resulting in successive years of co-authored publications.

The score of a facilitator is the sum of all the first-time collaborations that can be traced back to him. This score increases with the number of first-time collaborations within a given year. Another factor is the number of years these first-time collaborations recursively lead to new ones. In the extended PASNP Consortium network of 4,261 researchers, most only facilitated an average of 1.17 first-time collaborations. First-time collaborations, as a weighted edge between two co-authors, averaged only 1.14 papers for a given year. On average, first-time collaborations recursively led to only 0.5 years (one year being a recursive step) of new collaborations.

In contrast, the top 15 prime facilitators in the extended PASNP Consortium network were responsible for the vast majority of first-time collaborations. The top 15 prime facilitators each had attributed to them a range of 207 to 1,118 first-time collaborations. We hypothesize that most collaborative relationships do not continuously endure over time. The exception, reflected in the score of the top prime facilitators, is when a team of researchers primarily co-author with each other over successive years.

Table 1 Summary statistics of the Pan-Asian SNP Consortium Network.

Year	# Papers	# Edges	Average author-node degree	# First-time collaborations	Average # authors/paper
2004	114	3,522	1.7	N/A	7.7
2005	123	3,965	1.9	167	7.4
2006	129	11,264	5.3	332	9.9
2007	157	5,389	2.5	479	8.0
2008	210	10,490	4.9	811	9.6
2009	229	12,976	6.1	1,843	9.4
2010	204	9,420	4.4	1,495	8.2
2011	225	11,244	5.3	2,078	9.1

Basic statistics on country level inter-Asian facilitation in the PASNP Consortium network

We present in Fig. 4 the country-level aggregation of the co-authorship network of the PASNP Consortium as year-slice graphs. The extended co-authorship of the PASNP Consortium members spanned across 45 countries within and beyond Asia. We define Asia as including at least those countries that participated in the PASNP Consortium. This is based on how the members reflexively defined themselves as an Asian research network connecting 11 countries: China, India, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand, and the United States.[3]

We examined the role of the PASNP Consortium in catalyzing an emerging Asian Research Area. Developed and emerging countries in the network showed different patterns in the rate of publication from 2004 to 2011. Table 5 shows how emerging countries in genomics research significantly increased their publication rate to at least 1.5 times that of 2004. For example, China's publications rose from 25 in 2004 to 95 in 2011, averaging 56.1 publications per year. In contrast, developed countries such as Japan had a steady rate of publication close to their average for the time period.

Inter-Asian collaborations

In addition to holistically examining the PASNP Consortium's co-authorship network, we also define a few criteria to assist in comparing portions of the publications made by PASNP Consortium researchers, as shown in Fig. 5. International (as opposed to national) publications involve two or more countries. As we used the 92 researcher names and affiliations to find and select the publications to represent the PASNP Consortium, this entails that at least one author was a consortium researcher, regardless of which country he was at the time. Inter-Asian publications are a proper subset of international publications that involve two or more consortium countries, in which at least one author was a consortium researcher based at a consortium country. Inter-Consortium publications involve at least two of the 92 consortium researchers. Finally, inter-Asian within Consortium is the intersection of the previous two definitions; i.e., publications that are authored by at least

---

[3] The US bio-chip company, Affymetrix, supported the consortium with equipment, training, and manpower. The company also opened its first overseas manufacturing plant in Singapore.

Table 2 Top 15 PASNP Consortium researchers as ranked by (1) our RF-metric and (2) betweenness centrality.

Prime facilitator			Betweenness centrality		
Author	Country	Score	Author	Country	Score
Wei Huang*	CHN	3,682	Li Jin*	CHN/USA	0.67
Li Jin*	CHN/USA	1,905	Katsushi Tokunaga*	JPN	0.51
Xuejun Zhang	CHN	1,489	Xuejun Zhang	CHN	0.46
Ying Wang*	CHN	1,200	Takashi Gojobori*	JPN	0.33
Jieming Chen*	TWN	982	Hidetoshi Inoko*	JPN	0.29
Shuhua Xu*	CHN	687	Sumio Sugano*	JPN	0.25
Wentao Yuan*	CHN	666	Wei Huang*	CHN	0.25
Shunqiang Gao	CHN	664	Ying Wang*	CHN	0.23
Kaiyue Zhang	CHN	641	Haifeng Wang*	CHN	0.20
Yuhui Zhang	CHN	599	Edison Liu*	SGP	0.19
Zhenyu Lu	CHN	506	Yoshiyuki Suzuki*	JPN	0.16
Xiaoyan Xiong	CHN	466	Jun Ohashi*	JPN	0.15
Edison Liu*	SGP	451	Yuan-Tsong Chen*	TWN	0.13
Guoping Zhao*	CHN	450	Yoshiyuki Sakaki*	JPN	0.11
Yong Cui	CHN	399	Yixue Li	CHN	0.10

Asterisks denote PASNP Consortium members

Table 3 Components of RF-scores of top 15 PASNP Consortium researchers.

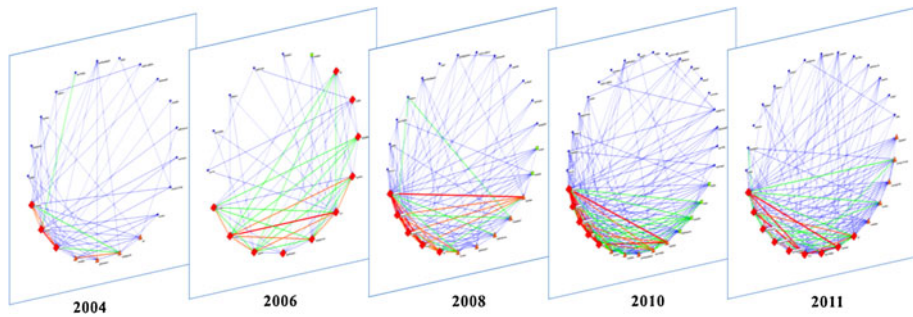| Facilitator C	Country	A. $RF_b(C)$	B. $RF_{\bar{b}}(C)$	C. $\frac{B}{A}$	# First-time collab. facilitated by C
Wei Huang*	CHN	3,682	1,299	0.35	1118
Li Jin*	CHN/USA	1,905	1,098	0.58	886
Xuejun Zhang	CHN	1,489	437	0.29	533
Ying Wang*	CHN	1,200	553	0.46	819
Jieming Chen*	TWN	982	300	0.31	304
Shuhua Xu*	CHN	687	129	0.19	175
Wentao Yuan*	CHN	666	142	0.21	214
Shunqiang Gao	CHN	664	65	0.097	49
Kaiyue Zhang	CHN	641	165	0.26	105
Yuhui Zhang	CHN	599	266	0.44	273
Zhenyu Lu	CHN	506	43	0.085	32
Xiaoyan Xiong	CHN	466	7	0.015	4
Edison Liu*	SGP	451	290	0.64	239
Guoping Zhao*	CHN	450	359	0.80	462
Yong Cui	CHN	399	190	0.48	207
Top 15 Average		986.5	365.2	0.37	361.3

Asterisks denote PASNP Consortium members

Fig. 4 The evolution of the PASNP network from 2004 through 2011. The vertices placed around the edge of the graph represent countries; and edges represent co-authorship magnitudes of varying strength (red high; green middle; blue low).

two consortium researchers, based at two or more consortium countries. Publications were filtered by title so that they would only count once.

From Fig. 6, we see that international publications were largely driven by inter-Asian publications, which comprised 82% of all the international publications. This reflects significant regional coherence as the 92 consortium researchers mostly co-authored with colleagues in Asia, both within and beyond the consortium.

Inter-Asian within consortium publications account for a surprisingly low percentage of the total number of publications during the entire period of analysis—2004 to 2011—at only 8.6% of the total publications. This rate is not much higher than before the start of the consortium in 2004. However, inter-consortium activity rose steadily since 2004, peaking at 96 papers published in 2009. This indicates that most inter-consortium publications occurred within each respective country rather than internationally. In other words, most collaboration fostered by the international PASNP Consortium occurred within national contexts.

Applying the RF-metric to examine country level facilitation in the PASNP Consortium network

We apply the RF-metric to examine inter-Asian facilitation as triadic closure among researchers at the country level. We then examine how researchers in China and Japan exhibited different patterns of research facilitation.
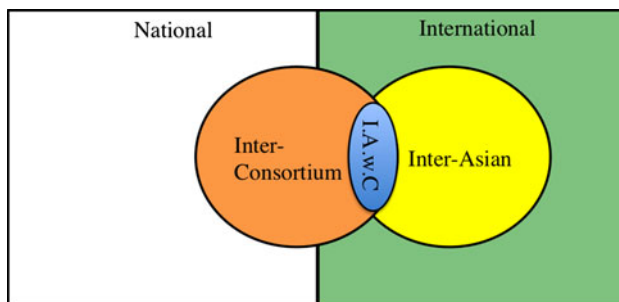

Fig. 5 Schematic showing the set-wise relationship between our defined selection criteria. Colors chosen to correspond to Fig. 6.
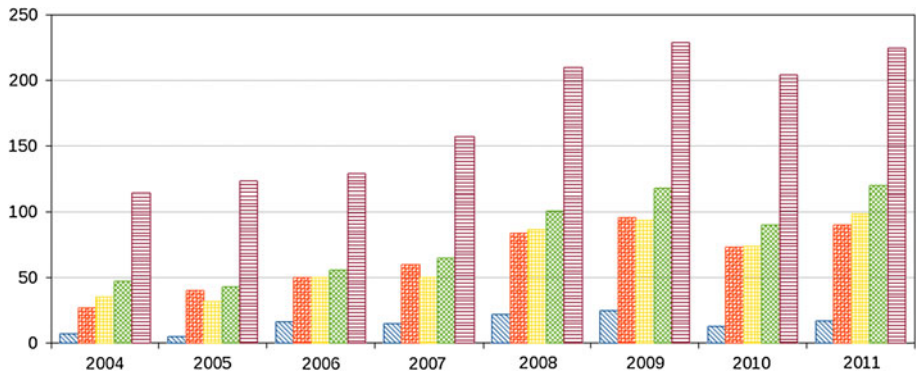
Fig. 6 Co-authored publications in the Pan-Asian SNP Consortium from 2004 to 2011.

Facilitated interaction within national contexts is most striking for Chinese authors. According to the prime facilitator metric, Chinese researchers made up the bulk of the top-ranked fifteen authors (Table 2) within the entire PASNP Consortium network who made connections leading to new collaborations. This was an unexpected empirical result, rather than a by-product of our metric.

What type of network structure would emerge if we used automatic methods to cluster researchers based on co-authorship rather than institutional and country affiliations? We employed the stochastic Chinese Whispers algorithm (Biemann 2006) to automatically group researchers in the network solely on the basis of the co-authorship graphs. The 4,261 co-authors of the extended PASNP Consortium network from 2004 to 2009 formed 178 clusters of varying sizes, shown in Fig. 7.[4] 50% of the clusters had less than 10 co-authors. Clusters having less than 50 members were evenly distributed with about 5 clusters for each size.

Twelve of these clusters were considerably larger than the rest and each dominated by a single country. These national clusters represented Japan, China, Korea, Singapore, India, Thailand, and the US. All had a minimum of 50 co-authors and, except for the Chinese cluster, showed at least 80% geographic homophily (McPherson et al. 2001). All the large clusters exhibit small-world network characteristics, as stipulated by Watts and Strogatz (1998).[5]

Examining why Chinese researchers were ranked highly, we discovered that the prime facilitator metric had detected a unique pattern of collaborative behavior. In the entire network, there were only four exceptions where a facilitated collaboration directly led to six successive years of new collaborations. The researchers involved in the cases were predominately from China (Table 4).

---

[4] The Chinese Whispers algorithm has been empirically shown to perform well in detecting small-worlds even in networks with skewed mixture of clusters and is faster than other network clustering algorithms such as minimum cut and hierarchical clustering which are capable of similar performances. Owing to the stochastic nature of the algorithm, it might not be possible to replicate the exact number of clusters on an arbitrary run on the dataset, yet the empirical results would still be the same on every run.

[5] Small-worlds are a family of networks that form a continuum extending between the class of regular networks and random networks. The mean geodesic distance and clustering coefficient are measures to quantify the small-worldliness of connected networks. Small-world networks typically have high clustering coefficients yet low mean geodesic distances.
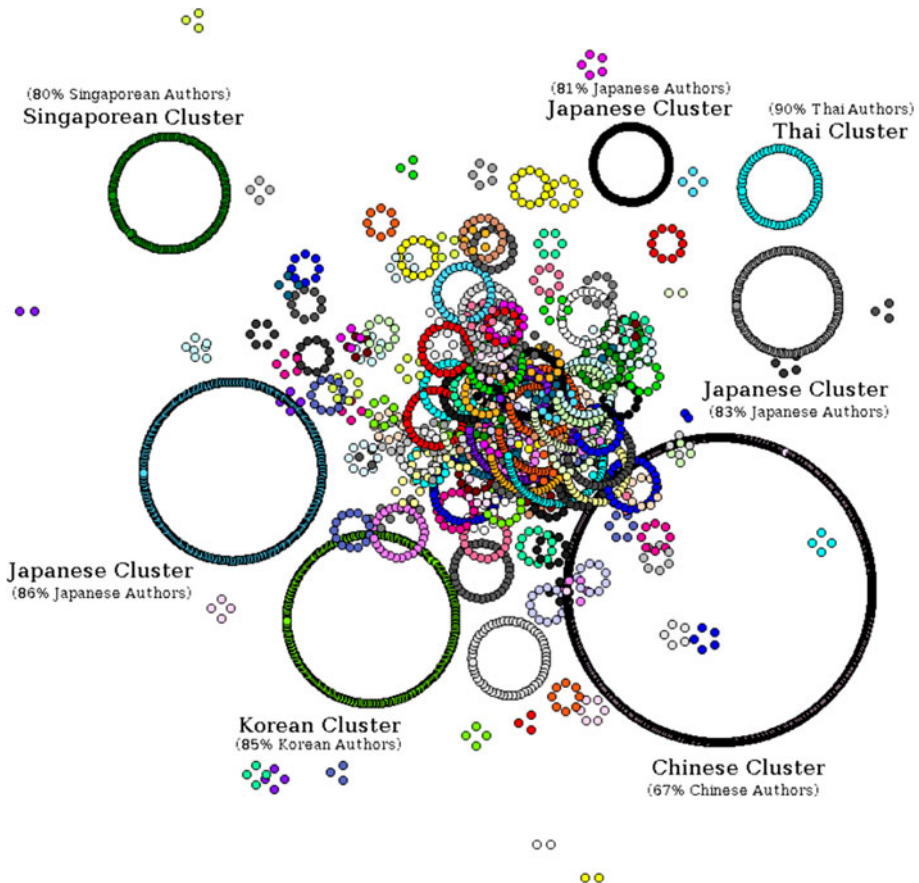
Fig. 7 Co-authorship clustering via the Chinese Whispers algorithm for the 4,261 researchers in the extended Pan-Asian SNP Consortium network.

Figure 8 shows one of these temporal chains of first-time collaborators involving the two highest-ranked prime facilitators, Wei Huang and Guoping Zhao.[6] In 2004, Xuejun Zhang had collaborated with both Wei Huang and Guoping Zhao. Xuejun Zhang became a prime facilitator of Wei Huang and Guoping Zhao when these two co-authored for the first time in 2005. In 2004, Wei Huang had also collaborated with Li Jin. In 2006, Wei Huang became a prime facilitator of the first-time collaboration between Guoping Zhao and Li Jin. This chain of new collaborations continued for six successive years. Above is one branch in the tree of facilitated first-time collaborations that began with Xuejun Zhang's facilitation of Wei Huang and Guoping Zhao. The full tree consists of 91 prime facilitators and 2,304 first time collaborations.

We examine how this micro-dynamic behavior of prime facilitators related to macro-dynamic patterns in international co-authorship. The co-authorship graph shows that the

---

[6] Note that no relevant activity occured in 2009.

Chinese cluster was exceptional with 900 co-authors, of which 67% were from China. In the entire network of 4,261 co-authors, the predominantly Chinese cluster contained the only four cases of first-time collaborations that led to six successive years of first-time collaborations. Table 4 and Fig. 8 show the six-year chain in which Xuejun Zhang facilitated collaboration between Guoping Zhao and Wei Huang.

This six-year chain formed a nearly complete graph (K_{44}) of repeat prime facilitators acting as a core group around which evolved a small-world. This explains how the highest scoring facilitators according to the RF-metric were able to generate successive recursive loops of new first-time relationships. According to the RF-metric, in order for a facilitator C to start a chain of first-time connections or earliest introductions $\overleftarrow{AB}(t)$, each dependent on the last, there must be new collaborators continuously added to the network. How such a continuous chain of new introductions can occur, emerged as an independent empirical result of the analysis on the Chinese cluster.

The core group consisted of 44 prime facilitators who made introductions at more than one time. These 44 repeat prime facilitators all belonged to a six-year chain and formed a nearly fully-connected network. This core group network and its corresponding Erdös–Rényi random graph with the same number of nodes and edges had exactly the same mean geodesic distance of 1.29 suggesting existence of a small-world. But, we do not find the network sparse enough to qualify as a small-world. According to Watts and Strogatz (1998), a network with n nodes and k edges per node should satisfy the inequality $n \gg k \gg ln(n) \gg 1$, to be considered for small-world analysis.

Clearly, the core group of 44 repeat prime facilitators had a higher degree of closeness than typical small-worlds. From 2004 to 2009, this core group facilitated first-time collaborations with a total of 153 new authors. These 153 non-facilitators formed a peripheral group around the core group. As we track the small-world parameters of this peripheral group through the years as shown in Fig. 9, we see that as the number of first

Table 4 Prime facilitators.

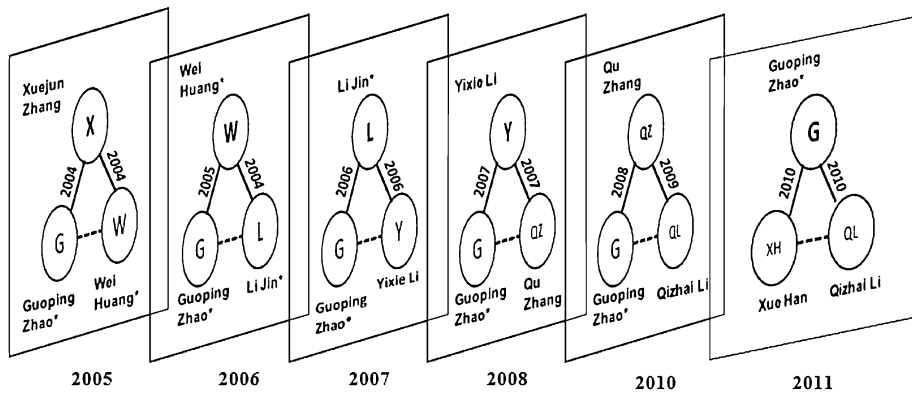| Year	Prime facilitators	First-time collaboration
2005	Xuejun Zhang; Zhenyu Lu; Xiaoyan Xiong	Guoping Zhao* (CHN)–Wei Huang* (CHN)
2005	Ying Wang*	Wei Huang* (CHN)–Yuhui Zhang (CHN)
2005	Kaiyue Zhang; Wei Huang*; Shunqiang Gao; Shuhua Xu*	Li Jin*(USA)–Ying Wang* (CHN)
2006	Edison Liu*	Jiangbo Liu (SGP)–Yixue Li (CHN)
2006	Wentao Yuan*; Wei Huang*	Yufeng Liu (TWN)–Ying Wang* (CHN)
2005	Wei Huang*; Shunqiang Gao	Ranajit Chakraborty (USA)–Ying Wang* (CHN)
2005	Wei Huang*; Shunqiang Gao	Bing Su (USA)–Ying Wang* (CHN)
2005	Pingping He; Wei Huang*; Yong Cui; Chengrang Li; Xuejun Zhang; Ming Li; Wentao Yuan*; Sen Yang;	Jieming Chen* (TWN)–Jiangbo Liu (SGP)
2006	Xuejun Zhang; Yuhui Zhang; Wei Huang*	Guoping Zhao* (CHN)–Li Jin* (USA)
2006	Min Gao; Jianjun Chen; Jieming Chen*	Wei Huang* (CHN)–Yuan–Tsong Chen* (TWN)

Asterisks denote PASNP Consortium members

Fig. 8 Chain of first-time collaborations over six successive years. Asterisks mark PASNP Consortium members. Dashed lines indicate facilitated relationships in the given year slice.

time collaborations start to rise after 2008, the mean geodesic distance of the network starts to fall.

The clustering coefficient, or cohesiveness, of this peripheral group gradually increased from 0.022 in 2004 to 0.14 in 2009. From 2009 onwards, this peripheral group with 153 nodes formed a small-world.

Table 6 shows that from 2009 to 2011, the mean geodesic distance of the peripheral group's network and its corresponding random graph is almost the same. The corresponding clustering coefficients differ by a large magnitude verifying small-worldliness. Thus, the core group reduced the degree of separation among the authors they facilitated, leading to the formation of a small-world.

This role of a core group of prime facilitators mediating the formation of a small-world around it is different from the role of hubs in a scale-free preferential attachment network (Barabási and Albert 1999). Preferential attachment models describe an actor's ability to attract more connections by virtue of the actor's popularity in the network. In a network with preferential attachment, the rate of additivity of new connections to an actor is governed by the richness of an actor's existing connectivity. They act as hubs to which other actors connect. Facilitation is an important factor in how networks grow and become more interconnected over time.

As most of the top research facilitators were from China, we examine the Chinese cluster as an example of an emerging country. For comparison in research facilitation, we then examine the Japanese clusters.

China: research facilitation in emerging countries

Based on our results, we choose China as an example of an emerging country (Fig. 7). From 2004 to 2011, China produced 449 publications from 757 researchers with an average of 3.9 publications per author. About one-third (248) of them facilitated 3,845 first-time collaborations with an average of 15.5 facilitations per author. These facilitations include both domestic and international collaborations.
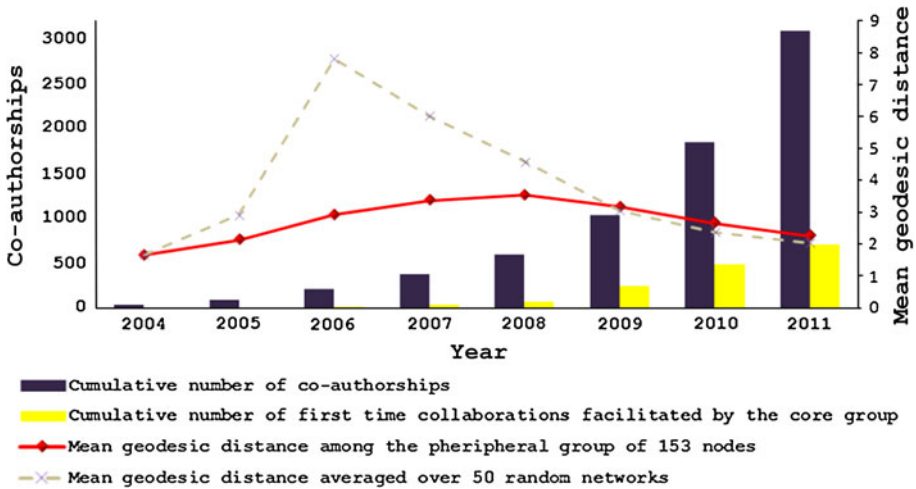
Fig. 9 Co-authorships and first-time collaborations from 2004 to 2011 among the 153 non-facilitators coincide with a drop in mean geodesic distance among the network of 153 non-facilitators for the same period. The mean geodesic distance was base-lined against the average mean geodesic distance of 50 random networks with the same number of nodes and edges.

Table 5 Distribution of publications among countries by year.

Group	Country	2004	2005	2006	2007	2008	2009	2010	2011	Avg # publications/ year
Emerging countries	China	25	19	30	40	70	96	74	95	56.1
	Singapore	18	20	17	18	35	34	32	54	28.5
	Korea	5	14	24	23	32	34	21	21	21.8
	Taiwan	4	4	19	12	26	21	40	36	20.2
	UK	5	0	5	15	16	21	13	17	11.5
	India	6	7	8	11	8	10	8	23	10.1
	Thailand	2	2	6	6	15	19	10	11	8.9
	Australia	2	2	0	1	4	9	3	14	4.4
	Indonesia	1	1	0	5	10	4	3	6	3.8
	Malaysia	1	0	0	0	0	3	5	10	2.4
Developed countries	Japan	64	71	67	74	90	75	68	76	73.1
	USA	33	26	35	35	53	62	47	43	41.8
	Germany	3	4	3	0	6	3	7	1	3.4
	Netherlands	2	1	0	4	2	7	4	2	2.8

Countries with an average annual publication of less than 2 are not shown in the table. This list includes Austria, Brazil, Canada, Colombia, Denmark, Estonia, Finland, France, Guinea, Hong Kong, Hungary, Iraq, Israel, Italy, Kenya, Lebanon, Mongolia, Norway, Pakistan, Philippines, Poland, Russia, Saudi Arabia, Spain, Sudan, Sweden, Turkey United Arab Emirates, Vietnam and Yemen.

Table 6 Small-world metrics for the peripheral group from 2009 onwards.

Year	C	L	C_{random}	L_{random}
2009	0.14	3.15	0.014	3.07
2010	0.22	2.66	0.04	2.36
2011	0.29	2.25	0.06	2.04

Variables C and L denote clustering coefficient and mean geodesic distance respectively.

Japan: research facilitation in developed countries

To analyze research facilitation in a developed country, we examine the two Japanese clusters (Fig. 7). For the same period, 1,546 Japanese researchers produced 585 publications with an average of 2.8 publications per author. Less than one-fifth (275) of them facilitated 2,097 first-time collaborations. On average, each author facilitated 7.6 first-time collaborations.

The marked difference in the number of first-time collaborations facilitated by the Japanese authors indicates that their collaboration network is structurally different from the Chinese. We note that the Chinese researchers facilitate first-time collaborations more than the Japanese as evident from Table 7. We also note that the Chinese network with an average degree of 42 is more densely connected than the Japanese network, which has an average degree of 23.3.

Table 7 Triad census of transitive triads and potentially transitive triads in Chinese co-authorship network vs Japanese co-authorship network.

Country	Year	# Potentially transitive triads (in thousands), α	# Transitive triads (in thousands), β	Percentage of β to α (%)
CHN	2004	0.61	0.00	0.00
	2005	4.01	0.15	3.79
	2006	30.39	0.28	0.91
	2007	20.22	0.54	2.67
	2008	56.11	0.60	1.07
	2009	266.14	2.39	0.90
	2010	281.59	4.67	1.66
	2011	452.58	2.94	0.65
	Total	1111.66	11.57	1.04
JPN	2004	2.99	0.00	0.00
	2005	18.72	0.18	0.97
	2006	45.54	0.27	0.59
	2007	52.99	0.29	0.56
	2008	181.48	0.76	0.42
	2009	164.91	0.51	0.31
	2010	191.22	0.75	0.39
	2011	299.95	1.36	0.45
	Total	957.80	4.13	0.43

The PASNP Consortium thus catalyzed a cascade of new first-time collaborations in China but less so in Japan.

Conclusion

Scientific collaboration has emerged as a key dimension for constructing the meta-geography of Asia. More than just physical topography, Asia has always been a conceptual framework whose imagined boundaries shift along with changing discourses on politics, economics, race, and culture (Lewis and Wigen 1997; Duara 2001). In the early 20th century, pan-Asian regionalism was a shared political ideology and identity against Western domination, uniting humanist intellectuals from China, Japan, and Korea to those as far as India, the Philippines and even Cuba (Karl 1998). By the end of the Cold War, rapidly developing tiger economies defined an exceptionalist Asia through greater economic regionalization, integration, and interdependence, embodied by pan-Asian organizational bodies such as APEC. At present, a pressing issue is whether scientific and technological collaboration is redefining Asia as an emerging research area.

There are a growing number of inter-Asian research consortiums, which like economic and political institutions, assert and thus reflexively define a pan-Asian regionalism. A largely Japanese led initiative backed by the Takeda Foundation has even called for the institutional formation of an Asian Research Area to parallel the European Research Area. They aim at fostering greater Asian scientific collaboration.

We find that Asia has emerged as an increasingly integrated research area through research facilitation, indirect, or mediated third-party introductions leading to triadic closures among co-authorship networks. We introduce a series of research facilitation or RF-Metrics to quantitatively measure the evolution of these relationships over time at the levels of individual authors, proximate groups, or recursive chains of introductions leading to new ones. As a case study, we applied our RF-Metrics to the development of the PASNP Consortium.

We built a software pipeline to gather our data and implement our RF-Metrics. The advantages of our system are its flexibility and availability. The publication records are collected from Google Scholar (GSNAP), and the author–affiliation metadata is extracted by Enlil. Although the paper only focuses on the PASNP network, our system can be applied to other scientific collaboration networks as well.

We treat the concept of Asia critically, as an historical social construction. The PASNP Consortium served as an ideal case study of a collaborative scientific network whose members reflexively asserted an Asian regional identity. Over the period of collaboration from 2004 to 2011, the consortium formed an increasingly dense network of newly formed connections and co-authored publications. However, the majority of collaboration within the consortium was among the members of the same country.

We also discovered that research facilitation played different roles for researchers of different statuses as well as developed and emerging countries. Principal investigators of institutions or consortiums ranked highly as research facilitators, but did not necessarily mediate connections that led to new ones. Leaders of close-knit research groups who collaborated frequently ranked highly and facilitated connections that led to successive chains of new first-time relationships.

Developed countries in the consortium had a steady rate of publication. In countries such as Japan, this rate was very high. Although the publication rate of Japanese researchers was high, these researchers did not rank well as research facilitators, indicating

that involvement in the PASNP Consortium did not catalyze many new connections. In contrast, emerging countries such as China not only increased their publication rate by more than 150%, but also closed nearly twice the ratio of potential triads compared to developed countries such as Japan.

We also found that a group of 44 mostly Chinese researchers, who collaborated frequently, facilitated first-time collaborations among a wider network of 153 researchers. This group of 44 frequent collaborators reduced the geodesic distance among the 153 researchers to form a peripheral small-world.

Research facilitation allows us to see how Asia as a research area has dynamically evolved in ways that are sometimes obscured by dyadic ratios of publication output. Examining transitive triads of individual authors enables us to see how relationships over time connected to each other at both micro- and macro-dynamic levels.

References

Barabási, A. L., & Albert, R. (1999) Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Basu, A. (2001). A comparative analysis of India and other Asian countries based on science, technology and development indicators. *Research Evaluation, 10*(1), 203.

Biemann, C. (2006). Chinese whispers An efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of the first workshop on graph based methods for natural language processing, TextGraphs-1* (pp. 73–80).

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior, 22*, 345–423.

Cho, P. S., Bullock, N., & Ali, D. (2013). The bioinformatic basis of Pan-Asianism. *East Asian Science, Technology and Society: An International Journal, 7*(2), 1–27. doi:10.1215/18752160-2142980.

Councill, I. G., Giles, C. L., & Kan, M. Y. (2008). Parscit: An open-source CRF reference string parsing package. In *Proceedings of the international conference on language resources and evaluation.* European Language Resources Association.

Do, H. H. N., Chandrasekaran, M. K., Cho, P. S., & Kan, M. Y. (2013). Extracting and matching authors and affiliations in scholarly documents. In *Joint conference on digital libraries.* New York: ACM (Forthcoming).

Duara, P. (2001). The discourse of civilization and Pan-Asianism. *Journal of World History, 12*(1), 99–130.

Frame, D. J., & Carpenter, M. P. (1979) International research collaboration. *Social Studies of Science, 9*(4), 481–497.

Garg, K. (2002). Scientometrics of laser research in India and China. *Scientometrics, 55*(1), 71–85.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology, 78*(6), 1360–1380.

Guan, J., & Ma, N. (2007). A bibliometric study of China's semiconductor literature compared with other major Asian countries. *Scientometrics, 70*(1), 107–124

Haustein, S., Tunger, D., Heinrichs, G., & Baelz, G. (2011). Reasons for and developments in international scientific collaboration: Does an Asia–Pacific research area exist from a bibliometric point of view? *Scientometrics, 86*(3), 729.

Karl, R. (1998). Creating Asia: China in the world at the beginning of the twentieth century. *The American Historical Review, 103*(4), 1096–1118.

Khan, G. F., & Park, H. W. (2012). Editorial: Triple helix and innovation in Asia using scientometrics, webometrics, and informetrics. *Scientometrics, 90*(1), 1.

Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science, 311*(5757), 88–90.

Kumar, R., Tripathi, R., & Tiwari, M. (2011). A case study of impact of patenting in the current developing economies in Asia. *Scientometrics, 88*(2), 575–587.

Leclerc, M., & Gagné, J. (1994). International scientific cooperation: The continentalization of science. *Scientometrics, 31*, 261–292.

Lewis, M. W., & Wigen, K. E. (1997) *The myth of continents: A critique of metageography*. Berkeley, CA: University of California Press.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values, 17*(1), 101–126.

Maoz, Z. (2009). The effects of strategic and economic interdependence on international conflict across levels of analysis. *American Journal of Political Science, 53*(1), 223–240.

McPherson, M., Smith-Lovin, L., & Cook J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444.

Murthy, R. (2009). India is "Thailand" to Asia, say scientists. *Asia Times Online*.

Normile, D. (2004). Consortium hopes to map human history in Asia. *Science, 306*, 1667.

Obstfeld, D. (2005) Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly, 50*(1), 100–130.

Okubo, Y., Miquel, J., Frigoletto, L., & Doré, J. (1992). Structure of international collaboration in science: Typology of countries through multivariate techniques using a link indicator. *Scientometrics, 25*, 321–351.

Rapoport, A. (1953). Spread of information through a population with socio-structural bias. *Bulletin of Mathematical Biology, 15*(4), 523–533.

Salton, G., & McGill, M. (1983) *Introduction to modern information retrieval*. Auckland: McGraw-Hill.

Schott, T. (1988) International influence in science: Beyond center and periphery. *Social Science Research, 17*(3), 219–238.

Schubert, A., & Braun, T. (1990) International collaboration in the sciences 1981–1985. *Scientometrics, 19*, 3–10.

Shils, E. (1975). Center and periphery. In *Essays in macrosociology* (pp. 3–16). Chicago: University of Chicago Press

Shin, J. C., Lee, S. J., & Kim, Y. (2012) Knowledge-based innovation and collaboration: A triple-helix approach in Saudi Arabia. *Scientometrics, 90*(1), 311.

Tang, J., Wu, S., Sun, J., & Su, H. (2012). Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '12)* (pp. 1285–1293).

The HUGO Pan-Asian SNP Consortium. (2009). Mapping human genetic diversity in Asia. *Science, 326*(5959), 1541–1545.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393*(6684), 440–442.

Wineman, J. D., Kabo, F. W., & Davis, G. F. (2009). Spatial and social networks in organizational innovation. *Environment and Behavior, 41*(3), 427–442.