Model for Dominating Authors

Theo thảo luận của anh Vũ và anh Hải về việc đưa trọng số đánh giá vai trò của tác giả bài báo tương ứng với vị trí của tác giả trong bài báo, Kiên có ý xây dựng mô hình như sau: (Ý tưởng này dựa trên ví dụ của anh Hải, và Kiên cũng đồng tình với ý tưởng này)

Giả sử một publication p có k authors theo thứ tự: $A_1, A_2, \ldots, A_k$. Khi đó chúng ta xác định một hàm trọng số $\beta_1, \beta_2, \ldots, \beta_k$ (phụ thuộc theo k) để đánh giá mức ảnh hưởng của p lên authors thõa mãn:

$$\beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_k \qquad \textit{and} \qquad \beta_1 + \beta_2 + \cdots + \beta_k = 1.$$

Khi đó

$$T(p, A_i) = W_{12}(p, A_i) = \beta_i \qquad \forall i = 1, \ldots, k.$$

Vấn đề chúng ta cần xác định $T(A_i, p) = W_{21}(A_i, p)$???

Xét một authors A có n bài báo $p_1, p_2, \ldots, p_n$. Dựa vào vị trí của A trong n bài báo này chúng ta xác định được ảnh hưởng của n bài báo này đến A như sau:

$$T(p_j, A) = \alpha_j \qquad \forall j = 1, \ldots, n.$$

Khi đó chúng ta không nên xem A có n bài báo mà nên xem A có

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n$$

bài báo. Từ đó ảnh hưởng của author A đến các bài báo được xác định như sau:

$$T(A, p_j) = W_{21}(A, p_j) = \frac{\alpha_j}{\alpha_1 + \alpha_2 + \cdots + \alpha_n} \qquad \forall j = 1, \ldots, n.$$

Chúng ta có thể xác định các trọng số $\beta_1, \beta_2, \ldots, \beta_k$ như ý tưởng anh Hải đưa ra.

Em viết sơ công thức tính để anh Hải cho chương trình chạy. Sau khi có kết quả, nếu tốt thì em sẽ viết công thức lại nghiêm túc hơn để chúng ta đưa vào bài báo. Anh Vũ và anh Hải xem rồi cho ý kiến nhé.

Hải comment: Anh đồng ý với Kiên về tổng các α_i, nhưng làm thế nào để so sánh và đánh giá các giá trị α_1,α_2...? Có cách nào tính được từ các β_j bằng việc chuyển vị như ví dụ của anh nêu ra không? Em Kiên.

Kiên comment: Các giá trị $\alpha_1, \alpha_2, \ldots, \alpha_n$ chính là các giá trị β đấy anh. Em viết như thế này để anh code chương trình dễ hơn, không thông qua ma trận. Em ví dụ nhé: Giả sử author A có 5 publications p_1, p_2, p_3, p_4, p_5.

- A đứng vị trí 1 trong publication p_1 và publication p_1 có 3 authors. Suy ra A có 1/2 (α_1) publication đối với p_1

(chúng ta xác định theo trọng số 1/2, 1/3 và 1/6)

– A đứng vị trí 2 trong publication p_2 và publication p_2 có 2 authors. Tương tự A sẽ có 1/3 (α_2) publication đối với p_2

(chúng ta xác định theo trọng số 2/3 và 1/3)

– A đứng vị trí 1 trong p_3 và p_3 có 1 author. A sẽ có 1 (α_3) publication đối với p_3.
– A đứng vị trí 1 trong p_4 và p_4 có 5 authors. A sẽ có $5/15 = 1/3$ (α_4) publication đối với p_4

(chúng ta xác định theo trọng số 5/15, 4/15, 3/15, 2/15 và 1/15)

– A đứng vị trí 3 trong p_5 và p_5 có 6 authors. A sẽ có 4/21 (α_5) publication đối với p_5

(chúng ta xác định theo trọng số 6/21, 5/21, 4/21, 3/21, 2/21 và 1/21)

Như vậy đối với 5 publications, theo cách tính chúng ta A sẽ có số bài báo là:

$$\alpha_1 + \alpha_2 + \ldots + \alpha_5 = \frac{1}{2} + \frac{1}{3} + 1 + \frac{1}{3} + \frac{4}{21} = \frac{33}{14}$$

Vậy điểm ảnh hưởng của A lên các publications và publications lên A sẽ là:

$$T(p_1, A) = \alpha_1 = \frac{1}{2} \qquad and \qquad T(A, p_1) = \frac{\alpha_1}{33/14} = \frac{7}{33}$$

$$T(p_2, A) = \alpha_2 = \frac{1}{3} \qquad and \qquad T(A, p_2) = \frac{\alpha_2}{33/14} = \frac{14}{99}$$

$$T(p_3, A) = \alpha_3 = 1 \qquad and \qquad T(A, p_3) = \frac{\alpha_3}{33/14} = \frac{14}{33}$$

$$T(p_4, A) = \alpha_4 = \frac{1}{3} \qquad and \qquad T(A, p_4) = \frac{\alpha_4}{33/14} = \frac{14}{99}$$

$$T(p_5, A) = \alpha_5 = \frac{4}{21} \qquad and \qquad T(A, p_5) = \frac{\alpha_5}{33/14} = \frac{4 \times 14}{33 \times 21}$$

Anh Hải xem có đúng ý anh chưa nhé.
Hi Kiên, như comment trong email: model này đã làm cho

$$W_{12} \ * W_{21} = I$$

Do đó, anh đề nghị cách tính W_{21} như sau:
Xét một authors A có n bài báo $p_1, p_2, \ldots, p_n$. Dựa vào thứ tự thời gian xuất bản và vị trí của A trong n bài báo này chúng ta xác định được ảnh hưởng của n bài báo này đến A theo quan điểm nếu bài báo được viết trước thì tác giả bỏ công sức bỏ ra nhiều hơn bài sau hay có ảnh hưởng lớn hơn, và như thế chúng ta có:

$$T(p_j, A) = \alpha_j * t_k \qquad \forall j = 1, \ldots, n. \forall k = 1, \ldots, n.$$

Khi đó chúng ta không nên xem A có n bài báo mà nên xem A có

$$\alpha_1 * t_1 + \alpha_2 * t_2 + \cdots + \alpha_n * t_n$$

bài báo. Từ đó ảnh hưởng của author A đến các bài báo được xác định như sau:

$$T(A, p_j) = W_{21}(A, p_j) = \frac{\alpha_j * t_j}{\alpha_1 * t_1 + \alpha_2 * t_2 + \cdots + \alpha_n * t_n} \qquad \forall j = 1, \ldots, n.$$

Các α_j xác định từ β_j như phần diễn giải của Kiên ở trên và t_k xác định theo việc đánh chỉ số năm xuất bản bài báo. Ví dụ : Giả sử author A có 5 publications p_1, p_2, p_3, p_4, p_5 với số năm xuất bản là $2011, 2008, 2003, 2009, 2008$ thì các trọng số t_k tương ứng là $1/15, 3.5/15, 5/15, 2/15, 3.5/15$)

Anh nghĩ thêm yếu tố thứ tự thời gian vào như thế này thì có gì đó chưa được hợp lý cho lắm nhưng chưa nghĩ ra cách nào hay hơn. Nhưng tạm thời sẽ tránh được việc tích nhân $W_12 * W_21 = I$ Và công thức cũng đã phức tạp lên nhiều rồi Rõ ràng nếu có 01 tác giả có xuất bản các bài báo cùng thời điểm thì công thức trở cũng chỉ về đơn vi đối với tác giả này mà thôi.

Nhờ Kiên và anh Vũ xem và cho thêm ý kiến nhé.