

**Handout 3.2: Outline of the Methodology for Tracking  
Link Formation in Online Data**

**Ref:** Chapter 4 of [EK] text.

**Note:** We present the methodology for tracking the formation of new links under **triadic closure**. This methodology was used in the 2006 study by Kossinets and Watts (reference [259] in the text).

**Steps of the Methodology:**

1. Consider two snapshots  $N_1$  and  $N_2$  of a social network at times  $t_1$  and  $t_2$  respectively, where  $t_1 < t_2$ .
2. For each value of  $k$ , let  $S_k$  denote the set of pairs of nodes  $\{x, y\}$  such that  $x$  and  $y$  have exactly  $k$  common neighbors in  $N_1$ , but the edge (link)  $\{x, y\}$  is *not* in  $N_1$ . (For some  $k$ , if  $S_k$  is empty, ignore set  $S_k$ .)
3. For each set  $S_k$  found in Step 2, let  $Q_k$  denote the subset of  $S_k$  such that for each pair  $\{x, y\}$  in  $Q_k$ , the edge (link)  $\{x, y\}$  *is* in  $N_2$ . For each value of  $k$ , compute the ratio  $T(k) = |Q_k|/|S_k|$ .

**Note:**  $T(k)$  is an empirical estimate of the probability that a link will form between two people who have exactly  $k$  common friends.

4. Plot  $T(k)$  against  $k$ . (We expect  $T(k)$  to increase with  $k$ .)

**Note:**  $T(0)$  represents the probability of link formation when two people have no common friend. So, a comparison of the value of  $T(0)$  with other values addresses the basic questions about triadic closure.