

# Birds Bring Flues? Mining Frequent and High Weighted Cliques from Birds Migration Networks

MingJie Tang<sup>1,3,\*</sup>, Weihang Wang<sup>1,3</sup>, Yexi Jiang<sup>5</sup>, Yuanchun Zhou<sup>1</sup>, Jinyan Li<sup>4</sup>,  
Peng Cui<sup>2,3</sup>, Ying Liu<sup>3</sup>, and Baoping Yan<sup>1</sup>

<sup>1</sup> Computer Network Information Center, Chinese Academy of Sciences

<sup>2</sup> Institute of Zoology, Chinese Academy of Sciences

<sup>3</sup> Graduate University of Chinese Academy of Sciences

<sup>4</sup> School of Computer Engineering, Nanyang Technological University

<sup>5</sup> School of Computer Science, Sichuan University, Chengdu

{tangrock, supercat0325, yexijiang}@gmail.com,  
zyc@.cnic.cn, jyli@ntu.edu.sg, cuipeng@ioz.ac.cn,  
yingliu@gucas.ac.cn, ybp@cnic.cn

**Abstract.** Recent advances in satellite tracking technologies can provide huge amount of data for biologists to understand continuous long movement patterns of wild bird species. In particular, highly correlated habitat areas are of great biological interests. Biologists can use this information to strive potential ways for controlling highly pathogenic avian influenza. We convert these biological problems into graph mining problems. Traditional models for frequent graph mining assign each vertex label with equal weight. However, the weight difference between vertexes can make strong impact on decision making by biologists. In this paper, by considering different weights of individual vertex in the graph, we develop a new algorithm, Helen, which focuses on identifying cliques with high weights. We introduce “graph-weighted support framework” to reduce clique candidates, and then filter out the low weighted cliques. We evaluate our algorithm on real life birds’ migration data sets, and show that graph mining can be very helpful for ecologists to discover unanticipated bird migration relationships.

**Keywords:** Graph Mining, Birds Migration, Birds Flues H5N1, Scientific data, Qinghai Lake.

## 1 Introduction

The Asian outbreak of highly pathogenic avian influenza H5N1 disease in poultry in 2003, 2004 and 2009 was unprecedented in its geographical extent. Its transmission to human beings showed an ominous sign of life-threatening infection [1]. Research findings indicate that the domestic ducks in southern China played critic role in virus reproduction and maintenance [9]. The major question is arising to understand the

---

\* This work was supported by Special Project of Informatization of Chinese Academy of Sciences in " the Eleventh Five-Year Plan", No.INFO-115-D02.

highly correlated species' habitat. It is critical for us to find the roots of the answers, such as: how the wild life and domestic poultry intersect together to translate virus to related places? How is the possibility of H5N1 that spilled over from the poultry sector into some wild bird species among habitats?

Clique is the most coherent and dense substructure among all kinds of subgraphs based on the assumption that there is at most one edge between any two vertices [3]. Discovering cliques from graph transaction database can provide insights about the underlying structure or relationships among different objects in graph transaction [5]. Meanwhile, researchers found that group of birds tended to move in ways that resembled weighted graphs, especially when a flock was active during their migration. For this matter, graph mining from birds' migration data does bring some aspiration for answering biological problem.

Traditional frequent cliques mining reflect information about the frequency of the presence or absence of a specified vertex label. However, in many cases it is possible that frequent cliques only contribute to a small portion of the overall "profit", whereas non-frequent cliques produce a large portion of the "profit". "Profit" could be defined as the object interest in different ways. In our study, biologists need to know active area of the birds by looking at the weights of the habitat combinations in order to consider the possibility of the bird interaction virus with poultry sector. The weight can be deemed as bird migration time on one habitat or density of bird satellite tracking location points. For instance, a clique,  $C_1$ , may be a frequent subgraph with frequency 60%, contributing 1% of the whole bird migration time. Another clique,  $C_2$ , may be a non-frequent clique with frequency 8%, contributing 20% of the whole migration time. Empirical experience from other ecological studies [10, 11] suggests that clique  $C_2$  would be much interest to biologists to track the bird flues transferring.

In this work, this new approach called *Helen* (it stands for **H**igh **w**Eighted **c**losed **cl**iqu**E** **m**i**N**ing) was proposed to mine high weighted closed clique from graph transaction databases. We first introduce graph-weighted support framework which adopts "downward close closure" to reduce the clique candidate sets. And then we prune the overestimated weighted cliques. The main contributions of our work are summarized as follows: (1) Convert the biological research problem associated with bird flues translation way into graph mining problem, (2) Present new birds' migration clique mining mode to find highly correlated habitat areas. (3) Provide important and hidden clues about the relationship of bird migration and H5N1 according to the results of our experiment.

The rest of this paper is organized as follows. Section 2 overviews the related work. Section 3 points out the desired results in which biologists interested and the practical challenges, then we introduce the preliminary concepts for discussion. Section 4 proposes the birds' migration clique mining model to find frequent and high weighted closed clique. Section 5 presents the experiment results and discuss the results for biological research ways. Finally, we summarize our works and point out the future research direction in the section 6.

## 2 Related Works

The applications of satellite tracking to bird migration studies have enabled considerable progress to be made with regard to elucidating the migration routes and stay sites

of various migratory bird species, with important implications, for example, for conservation [11]. Our previous works [2] use clustering and association rules to discover bird migration habitats, site connectedness and migration routes. However, Biologists found that bird migration routes in small range of area usually are graph patterns rather than simple sequence.

In the previous published literature, we are unable to find any work on discovering weighted cliques from graph database, however, lots of works deal with mining the frequent clique and quasi-clique from multiple graphs. Pei et al. [5] proposed an algorithm called Crochet to mine cross quasi-cliques from a set of graphs. Later, Wang et al. [3] studied the problem of mining frequent closed cliques from graph databases. We adopt their clique enumeration and pruning idea to find frequent clique (in the section 4), and extend this by using the graph-weighted and rechecking to get high weighted closed cliques (see section 4.2). Later, Zeng et al. [4] studied a more general problem formulation, that is, mining frequent closed quasi-cliques from graph databases, and proposed an efficient algorithm called Cocain to solve the problem.

Meanwhile, an effort was done on assessing weighted association rules mining in the last decade using either the average weight value of the items comprising this itemset, or utilizing weighted framework to evaluate the weighed association rules [7]. Presumably the most relevant work to our current study was done by Liu [6], this paper proposed two phase algorithm to deal with “utility mining” problem. We use their two phase principle, but instead of using utility principle we use weighted-support framework which is easier for interpretation (see section 3.2)

### 3 Problem Formulation

#### 3.1 Desired Results and Challenges

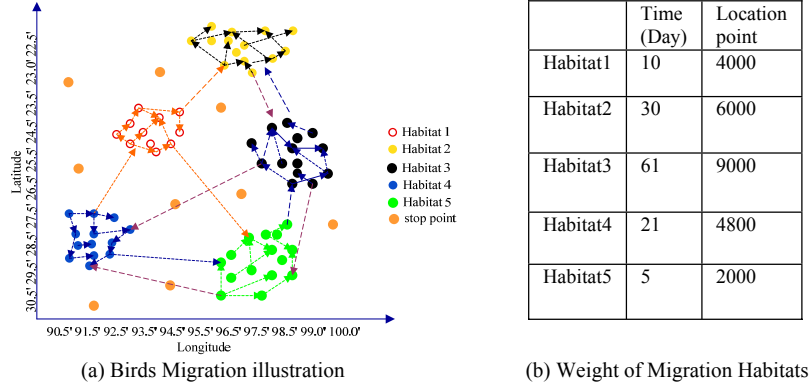
Supposed in the Figure 1(a), migration routes of birds can be regarded as one kind of graph, where habitat can be deemed as vertex nodes and the migration routes can be treated as edges. And birds’ migration routes can be made of one graph database. Analyzing cliques from this graph database would give important knowledge about the possibility of birds spillover H5N1 among habitats in the same clique. There are two kinds of cliques that we hope to discover: frequent clique and high weighted clique.

Frequent Clique mining is such a process: given a graph transaction database  $D$  and a minimum support threshold  $\min\_sup$ , identify the complete set of cliques in  $D$  that are both frequent and closed. Several frequent cliques mining methods [3, 4, 5] could be deployed to solve frequent clique mining.

In some scenarios, high weighted cliques can provide useful information if we pay attention to the graph vertex weight. For example, in Figure 1 (a) and its related table Figure 1(b), several factors such as the number of migration points or the time spending on a particular habitat can be deemed as weight. If we dismiss this kind of information which would influence biologists to judge the possibility of birds transfer avian virus to domestic ducks or other birds [9], the mining results would not be “interest”.

#### 3.2 Problem Definition

We start with the introduction of a set of terms that leads to the formal definition of high weighted cliques mining problem. The same terms are given in [3,4,5].



**Fig. 1.** Bird migration Space defined by Longitude and Latitude in the left. Point with different color and lines stand for tracked bird’s migration coordinates and migration routes, separately. In the right table is birds’ active information associate with their migration habitat.

**Notations Description**

- $V: V = \{v_1; v_2; \dots; v_k\}$ , the set of vertices
- $E: E \subseteq V \times V$ , the set of edges
- $L$ : the set of vertex labels
- $F: F: V \Rightarrow L$ , the mapping function from labels to vertices.
- $G: G = (V; E; L; F)$ , an undirected vertex-labeled graph
- $|G|: |G| = |V|$ , the cardinality of  $G$
- $G(S): G(S)$ : the induced subgraph on  $S$  from  $G, S \subseteq V(G)$

In this paper, we consider simple graph only, which does not contain self-loops, multi-edges, and edge labels. A clique is a fully connected graph and each pair of vertices in  $V$  there exist an edge in  $E$ . The size of a clique is defined by the number of vertices it contains, i.e.,  $|V|$ . A clique with  $n$  vertices is called an  $n$ -clique and the number of edges in the  $n$ -clique is  $n*(n-1)/2$ . For instance, given the graph database in Figure 2 and  $min\_support=2$ , frequent 3-clique  $G_1$  and 4-clique  $G_2$  are illustrated at the right side. We use canonical code to present each clique, canonical code representation is defined as the minimum string among all its possible strings such as [3]. For example, Graph 3 in Figure 2(b) is a clique with 4 vertexes, the canonical form CF is represented as the string “ABDE”, which is the smallest string with the combination of four letters “A”, “B”, “D” and “E”. In the rest of paper, clique will be mentioned by their canonical form directly. Depend on the canonical form of clique; the sub-clique relationship can be changed to subsequence relationship. If clique  $C_1$  and  $C_2$  with canonical form  $CF_1$  and  $CF_2$  respectively,  $C_1 \subseteq C_2$  iff  $CF_1$  is a sub string of  $CF_2$ .

Frequent clique mining has been introduced in [3,4]. In the rest of paper, weights for the graphs vertex are considered, and high weighted clique mining are discussed mainly. Before discussion, some important definitions are given below.

**DEFINITION 2.1 (Vertex weighted):** The weight value  $w(L)$  means the significance for certain vertex label  $L$ . A graph is a set of weighted labels, each of which may

appear in multiple graphs with same weight. For instance, in the Table 2, weight of the graph vertex in the Figure 2 are described:  $w(A)=7$  ,  $w(B)=6$  ...  $\square$

**DEFINITION 2.2 (Weighted graph):**  $G_w = (V;E; L; F;W)$ , an undirected vertex-labeled Weighted graph. The weighted graph is only considered with the weighted for the vertices label.  $\square$

**DEFINITION 2.3 (Weight of graph):** Weight ( $G^w$ ) is the one graph weight. It could be sum up weight of all vertex labels in one graph simply.

$$\text{Weight}(G) = \sum_{i=1}^{|V|} w_i \tag{1}$$

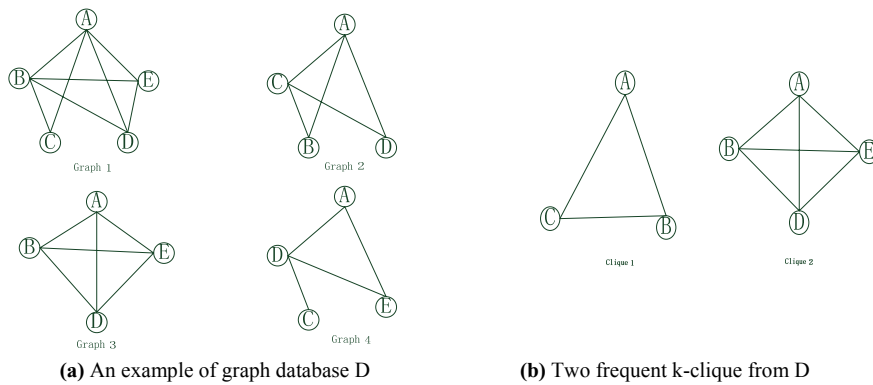
Where the  $|V|$  is the number of vertex label, the  $w_i$  is the vertex weight. For example, in the Figure 2 (a) and related weight Table 1:

$$\text{Weight}(G1) = \{w(A)+w(B)+w(C)+w(D)+w(E)\} = (7+6+2+14+20) = 49$$

**DEFINITION 2.4 (High Weighted Clique):** The high weighted clique can be defined as the sum of the weight of the graph and the weight of the fraction of transactions that the graph occurs in. Thus, one clique is high weighted clique if:

$$\text{WSP}(C) = \frac{\sum_{C \subseteq G_i} \text{Weight}(C)}{\sum_{i=1}^{|D|} \text{Weight}(G_i)} > \epsilon \tag{2}$$

Where the  $\text{Weight}(G)$  and  $\text{Weight}(C)$  is the weight of one graph defined in the DEFINITION 2.3.  $|D|$  is the number of graphs in one graph database.  $\epsilon$  is viewed as user's interest. In addition, one clique  $C$  is a **High Weighted Closed Clique (HWCC)**: If there does not exist another clique  $C'$  such that  $C \subseteq C'$ ,  $\text{WSP}(C') = \text{WSP}(C)$  and  $\text{WSP}(C) > \epsilon$ .  $\square$



**Fig. 2.** A graph database and parts of its sub graph

**Table 1.** The weight table for each vertical label in the Figure 2 (a)

Label	A	B	C	D	E
Weight	7	6	2	14	20

**Problem Statement:** Given one Graph database  $D$  and the related vertex weighted table  $WT$ , weighted clique mining is to find all high weighted closed clique. For example, in Figure 2 and Table 1,  $WSP(\text{Clique 1}) = WSP(ABC) = (15 \times 2) / (49 + 29 + 47 + 43) = 0.17$ ,  $WSP(\text{Clique 2}) = WSP(ABDE) = (47 \times 2) / (49 + 29 + 47 + 43) = 0.55$ . If  $\mathcal{E} = 0.5$ , clique(ABC) is a low weighted clique and should not be considered.

The graph database in the Figure 2(a), and graph database related weighted table (Table 1) will be our running examples in the rest of our paper.

#### 4 Birds Migration Closed Clique Mining Model

We utilized a data mining framework to discover the frequent and high weighted cliques as in the Figure.3. A clustering algorithm in [2] is developed in this system to find sub-areas with a dense location points relative to the entire area. Then we adopted clique mining to discover the frequent and high weighted clique between the discovered habitats. Because of space limitation, details of frequent closed clique mining approach CLAN can be reached from paper [3].

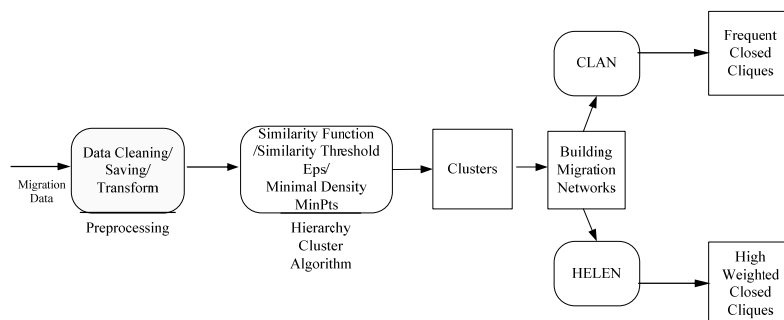


Fig. 3. System Framework of Cliques Mining Toward Birds Migration

##### 4.1 HELEN: High Weighted Closed Cliques Mining

Motivated by discovering high weighted cliques from birds' migration graph, we intend to extend traditional frequent graph mining method to meet our requirements. However, the "downward closure property" in Apriori-based approach cannot apply to the weighted clique mining directly due to weight bias support. Considering the challenges, "high graph-weighted" that owns downward closure property is deployed to reduce clique candidates. In the second place, rechecking is explored to filter out the high graph-weighted cliques that are indeed low high weighted clique. This approach was called HELEN.

**4.1.1 High Graph-Weighted Support Framework to Prune Candidates**

**DEFINITION 4.1. Graph-Weighted Clique:** The graph-weighted of a clique C, denoted as tw(C), is the sum of the graph weight of all the graph that C embedded in:

$$tw(C) = \sum_{C \subseteq G_w \subseteq D} Weight(G_w)$$

For example clique(ABC) in Figure 2, tw(clique ABC) = (W(G1)+W(G2))=(49+29)=78.

**DEFINITION 4.2. High Graph-Weighted Clique:** For a given clique C, C is a high graph-weighted supported clique if

$$GWSP(C) = \{tw(C) / (\sum_{i=1}^{|D|} Weight(G_i))\} > \epsilon',$$

**Theorem 1. The graph-weighted Downward Closure Property** indicates that when K-1 Clique is not high graph-weighted significant clique, the K clique can never be the high graph-weighted significant clique as well. In the process of enumerating clique, only high graph-weighted (K-1)-clique can be added as the candidates to extend K-clique.

**Proof:** Let  $T(K)$  be the collection of the graph transactions containing K-Clique and  $T(K-1)$  be the collection of transactions containing (K-1)-Clique. Although (K-1)Clique  $\subseteq$  K-Clique, the graph support K-clique will decrease as the K increasing, thus the  $T(K-1)$  is a superset of  $T(K)$ .

$$tw((K-1)clique) = \sum_{(k-1)clique \subseteq G_m} Weight(G_m) > tw((K)clique) = \sum_{(k)clique \subseteq G_n} Weight(G_n)$$

**Then:**

$$\{tw((K-1)clique) / \sum_{i=1}^{|D|} Weight(G_i)\} > \{tw((K)clique) / \sum_{i=1}^{|D|} Weight(G_i)\} > \epsilon'$$

**Theorem 2.** Let HGWCC be the collection of all high graph-weighted closed cliques in a transaction database D, and HWCC be the collection of high weighted closed cliques in D. if  $\epsilon' = \epsilon$ , HWCC  $\subseteq$  HGWCC

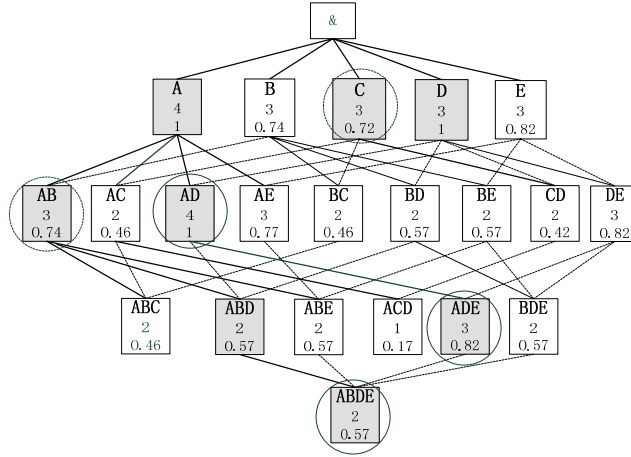
**Proof**

$\forall C \subseteq HWCC$ , if C is a high weighted clique, and  $C \subseteq HGWCC$

$$\epsilon' = \epsilon \leq WSP(C) = \frac{\sum_{C \subseteq G_i} Weight(C)}{\sum_{i=1}^{|D|} Weight(G_i)} < \frac{\sum_{C \subseteq G_i} Weight(G_i)}{\sum_{i=1}^{|D|} Weight(G_i)} = GWSP(C)$$

Thus, C is high graph-weighted clique and  $C \subseteq HGWCC$ .

To illustrated the process of clique enumeration: one lattice in the Figure.4 is built from the graph database in our running example (in section 3.2).The sub-clique relationship between two cliques can be represented by their canonical forms and they can be conceptually organized into lattice like structure in depth search first order. Each box represents one clique and its canonical form, edge between two boxes



**Fig. 4.** Clique traverse-lattice tree related to the example in section 3.2. Canonical form boxes covered by circle (solid and dashed) are high graph-weighted closed cliques when Weighted Support  $\mathcal{E}$  is 0.5. Gray-shaded boxes denote the search space. The number in the middle and bottom of canonical box is occurrence and graph-weighted support, separately. Clique with solid circle are the high weighted cliques after pruned.

means the sub-clique relationship. One such example in the lattice tree is one of cliques: ABE. It is supported by graph 1 and 3, and its frequent support is 2. Its graph-weighted support is  $(49+47)/(49+29+47+43)=0.82$ .

**4.1.2 Rechecking Procedure**

From phase 1 one have generated a set of High Graph-Weighted Closed Clique (HGWCC), but it may contain some false-positive results, we call such sub-clique **Pseudo High Weighted Closed Clique (PHWCC)**. We calculate its Weighted Support for each HGWCC. If it is a PHWCC, it should be pruned, otherwise it is kept. For example, in the Figure.4 Clique(AB) with graph-weighted support 0.74 should be pruned, because its weighted support is 0.29 and is lower to  $\epsilon=0.5$ . What is more, in order to reduce the computation cost, the weight of the entire graph and the result clique should be calculated beforehand and stored in a data structure. Since the number of graph and result clique is not huge, and the weight of each graph or clique is represented as a single value, the space cost is not high.

**5 Experiments**

In this section, we present empirical results. At first, the real application of birds' migration database is introduced. The birds migration location data sets are converted into a graph database contain 59 graphs. On average, there are 314 edges and 37 vertices in each graph. The maxim one is the one graph from bar headed goose data sets, which owns 540 edges and 67 vertices. Moreover the algorithm CLAN and HELEN is discussed how to discover some highly related birds' migration habitats, and present some empirical results in the section 5.1. In addition, we would evaluate the HELN



algorithm's efficiency and scalability (see section 5.2). Finally, we discuss the relationship between highly closed birds' migration habitats with H5N1 incidents in 5.3. The experiments are performed on a 1.83 GHZ Inter(R) core(TM) CPU with 2G memory and Windows XP platform. The program is implemented in Java. All of our results are embedded into the Google Map.

### 5.1 Frequent and High Weighted Closed Cliques from Birds Migration Network

Table 2 shows the number of HGWCC and HWCC. As the decrease in the closed large clique would bring an increase on the small closed cliques for compensate, the total amounts of high weighted closed clique will not change totally as a result. In order to evaluate if the high weighted mining results is useful to the biologist research, we compare the high weighted cliques with the frequent cliques mined by traditional methods. We do observe a number of interesting cliques. For example, a clique in Fig.6(a) is a not frequent item (its frequent is 3), however, its contribution the total time of birds spring migration is more than 5.2%.

**Table 2.** Experiment summary of birds' migration data

Minimum Weighted support Threshold	Run Times (Seconds)	#Candidates Cliques (Size>2)	#High Graph-weighted closed clique	#High Weighted closed clique
0.5	20	27	15	8
0.4	50	56	34	14
0.3	109	94	53	31
0.2	172	122	79	44
0.1	694	145	91	70

### 5.2 Efficiency and Scalability Test of HELEN

To evaluate the efficiency of the algorithm, the proposed algorithm HELEN was compared with CLAN. The desired results sets of HELEN and CLAN are different, and their performance is not suitable for comparable. However, our purpose here is show that our algorithm could handle high weighted graph mining problem without increase time consuming greatly. At first, we compare two algorithms with birds' migration data sets. Figure.5(a) shows the runtime of our algorithm by varying the weighted support threshold from 0.1 to 0.5. The results illustrated that since the number of candidate cliques decreases as the minimum weighted support increases, the execution time decreases, correspondingly. Time cost of rechecking would not increase greatly comparing to CLAN since we have saved parts of results before (see section 4.1.2). We also evaluate HELEN's scalability using several real databases in terms of the base size. In Figure.5(b) we replicated the birds' migration graphs from 2 to 16 times. It is evident that HELEN shows a linear scalability in runtime against the number of graphs in the database.

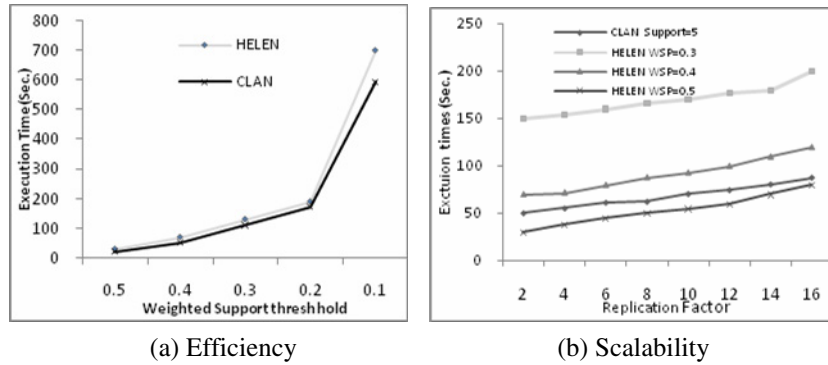
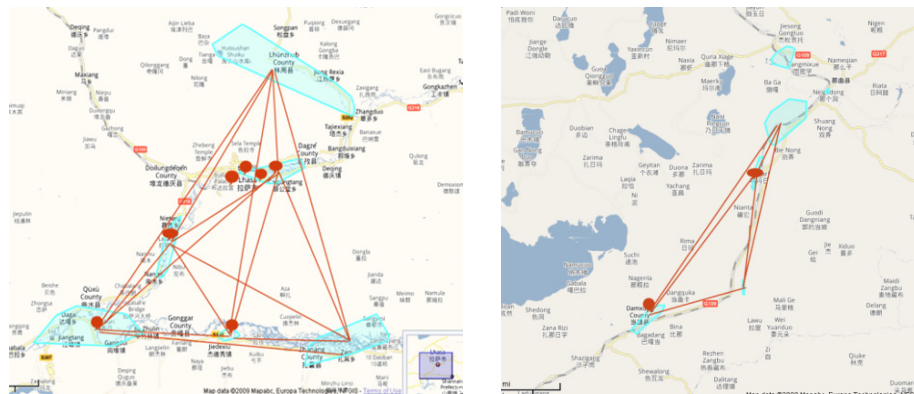


Fig. 5. Efficiency and Scalability Test of HELEN

### 5.3 Waterbirds Movements in Relation to High Related Habitats and H5N1 Outbreaks

Information about H5N1 outbreaks were obtained from the Ministry of Agriculture of the People’s Republic of China Database and OIE Database for the period 16 February 2004–18 May 2009. From our experiment results in the Figure 6, the correlation between birds’ migration action and the timing of H5N1 incidents in waterbirds region around Lasha, China was very high. This place is one of the most important areas for waterbirds to overwinter, and with high density of population and poultry. The high weighted clique in the Figure 6 means that waterbirds incline to stay those habitats in a longer time, since both of cliques have high weighted support 5.2% and



(a) One clique with frequent 3 and the Weighted support is 5.2% (b) One clique with frequent 2 and the Weighted support is 3.1%

Fig. 6. High Weighted Cliques related to Birds Migration Habitats and H5N1 outbreak locations including wild and domestic birds. Circle in blue (habitats), line in red (migration routes), and diamond circle (H5N1 once outbreak location). Vertex weight is time of birds’ spring migration, which lasts from 2008 Oct 10th to 2008 Nov 21th.

3.1%. For this matter, we can see that H5N1 outbreaks involving waterbirds occurred during winter, when the potential for interaction with poultry and probability of direct transmission from poultry to migratory waterbirds was predicted to be highest. The highly closed habitats also could be consider follow: waterbirds are directly infected from poultry (i.e., spillover), and they may be responsible for local movement of virus regionally, followed by the potential to transmit virus back to poultry (i.e., spillback)[10,11].

## 6 Conclusion

In this paper, we suggest to explore the field by using the location data information as a supplement data mining process which can provide an alternative approach for traditional bird telemetry data analysis: visual observation from the location points. In order to discover high weighted cliques, we develop new algorithm HELEN. Our experiment shows that frequent and high weighted clique mining do provide an effective assistance for biologists to discover new correlated relationship between habitats. In the future, we plan to extend our current work to address several unresolved issues. Specifically, we intend to extend the techniques proposed in this paper to mine high weighted closed quasi-cliques.

## References

1. Liu, J., et al.: Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science* 309, 1206 (2005)
2. Tang, M., Zhou, Y., Cui, P., Wang, W., Li, J., Zhang, H., Hou, Y., Yan, B.: Discovery of Migration Habitats and Routes of Wild Bird Species by Clustering and Association Analysis. In: Huang, R., Yang, Q., Pei, J., Gama, J., Meng, X., Li, X. (eds.) ADMA2009. LNCS, vol. 5678, pp. 288–301. Springer, Heidelberg (2009)
3. Wang, J., Zeng, Z., Zhou, L.: CLAN: An Algorithm for Mining Closed Cliques From Large Dense Graph Databases. In: ICDE 2006 (2006)
4. Zeng, Z., Wang, J., Zhou, L., Karypis, G.: Coherent closed quasi-clique discovery from large dense graph databases. In: SIGKDD 2006 (2006)
5. Pei, J., et al.: Mining Cross-graph Quasi-cliques in Gene Expression and Protein Interaction Data. In: ICDE 2005 (2005)
6. Ying, L., Liao, Choudhary, A.: A Fast High Utility Itemsets Mining Algorithm. In: UBDM 2005 (2005)
7. Tao, F., et al.: Weighted Association Rule Mining using Weighted Support and Significance Framework. In: SIGKDD 2003 (2003)
8. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In: Proc. of the 20th VLDB Conference (1994)
9. Li, et al.: Numbers and distribution of waterbirds and wetlands in the Asia-Pacific region: results of the Asian Waterbird Census Wetlands International (2007)
10. Newman, S.H., Iverson, S.A., et al.: Migration of Whooper Swans and Outbreaks of Highly Pathogenic Avian Influenza H5N1 Virus in Eastern Asia. *PLoS ONE* 4(5) (May 2009)
11. Sturm-Ramirez, K.M., Hulse-Post, D.J., Govorkova, E.A., Humberd, J., Seiler, P., et al.: Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia? *J. Virol.* 79, 11269–11279 (2005)