CSE 515 Multimedia and Web Databases

Phase #1

(Due Sept 21st 2014, midnight)

Description: In this project, you will experiment with

- epidemic simulation data sets and
- time series.

This project phase will be performed by each group member; but, you will get group grades.

- Download the sample data files from the project directory. See also the attached data description file.
- **Task 1:** Implement a program which creates an *epidemic words* dictionary: Given a directory, *dir*, window length, w, a shift length, h, and a resolution, r,
  1. for each data file, $f \in dir$, in the given directory, the program
     (a) normalizes the data file to values between 0.0 and 1.0,
     (b) quantizes the entries into r levels by dividing the range 0 to 1 into r Gaussian bands (with parameters: mean $\mu = 0$ and standard deviation $\sigma = 0.25$)
     – the length of the i^{th} band is equal to

$$length_i = \frac{\int_{(i-1)/r}^{i/r} Gaussian_{(\mu=0.0,\sigma=0.25)}(x)\ \ \delta x}{\int_0^1 Gaussian_{(\mu=0.0,\sigma=0.25)}(x)\ \ \delta x}$$

     – for each level, the center of the band is used as its representative.
     (c) for each state, s, the program
     i. moves a w-length window on the corresponding time series (by shifting it h time units at a time), and
     ii. for each window, starting at t, writes the pair $\langle idx, \vec{win} \rangle$, where
     – $idx = \langle f, s, t \rangle$ and
     – $\vec{win}$ is quantized content of the w-length window starting at time t,
     into a given file, $epidemic\_word\_file$.
- **Task 2:** Implement a program which reads a connectivity graph, G, a file, $epidemic\_word\_file$. Given a weight, $0 \leq \alpha \leq 1$, the program does the following: for each pair

$$w_i = \langle idx_i, \vec{win}_i \rangle,$$

the program

1. writes the pair
$$w_{avg,i} = \langle idx_i, \vec{win}_{avg,i} \rangle,$$
where
$$\vec{win}_{avg,i} = \Big(\alpha \times \vec{win}_i\Big) + \Bigg((1-\alpha) \times AVG\Big\{\vec{win}_j \mid \big(s_j \in 1HN(G, idx_i.s)\big) \wedge \big(idx_j = \langle idx_i.f, s_j, idx_i.t\rangle\big)\Big\}\Bigg)$$
into the file *epidemic_word_file_avg*.
2. writes the pair
$$w_{diff,i} = \langle idx_i, \vec{win}_{diff,i} \rangle,$$
where
$$\vec{win}_{diff,i} = \Bigg(\vec{win}_i - \Bigg(AVG\Big\{\vec{win}_j \mid \big(s_j \in 1HN(G, idx_i.s)\big) \wedge \big(idx_j = \langle idx_i.f, s_j, idx_i.t\rangle\big)\Big\}\Bigg)\Bigg) \div \vec{win}_i$$
and where $\div$ denotes element-wise division operation, into the file *epidemic_word_file_diff*.

Above, $1HN(G, s)$ denotes the set of 1-hop neighbors of the state s according to the connectivity matrix G.

- **Task 3:** Given a word $\langle idx, \vec{win} \rangle$, let the length (norm-2), $|\vec{win}|$, of the vector, $\vec{win}$, denote the strength of the word. Implement a program which lets the user
  - select an epidemic simulation file, f, and view it in the form of a heat map, and highlight (based on the user's choice)
    - the two windows, corresponding to f, with the highest and lowest strengths in the file *epidemic_word_file*,
    - the two windows, corresponding to f, with the highest and lowest strengths in the file *epidemic_word_file_avg*,
    - the two windows, corresponding to f, with the highest and lowest strengths in the file *epidemic_word_file_diff*.

    For each window, $\langle idx, \vec{win} \rangle$, please visualize the state $idx.s$ and the states in the 1-hop neighborhood of $idx.s$.

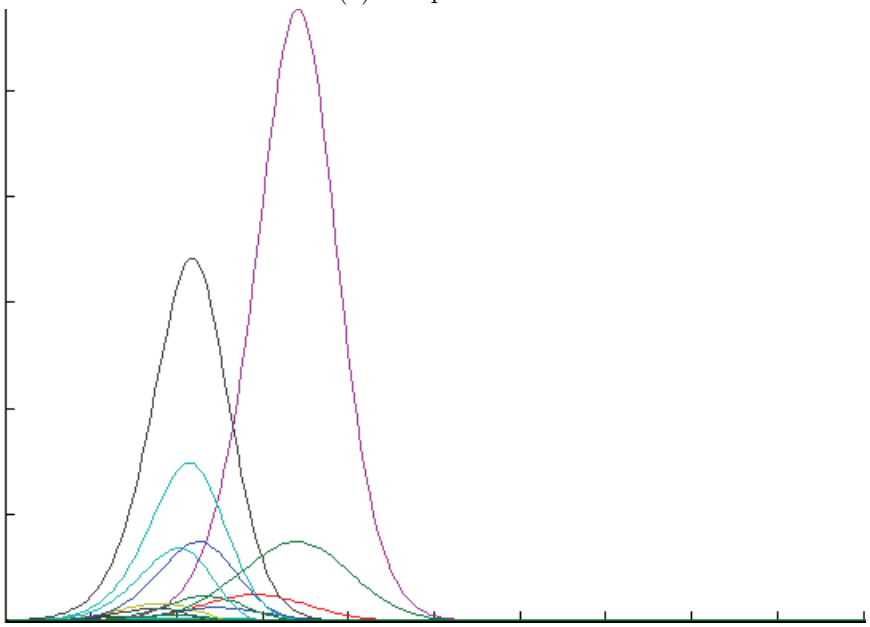
Deliverables:

- Your code (properly commented) and a README file.
- Your outputs for the provided sample inputs.
- A short report describing your work and the results.

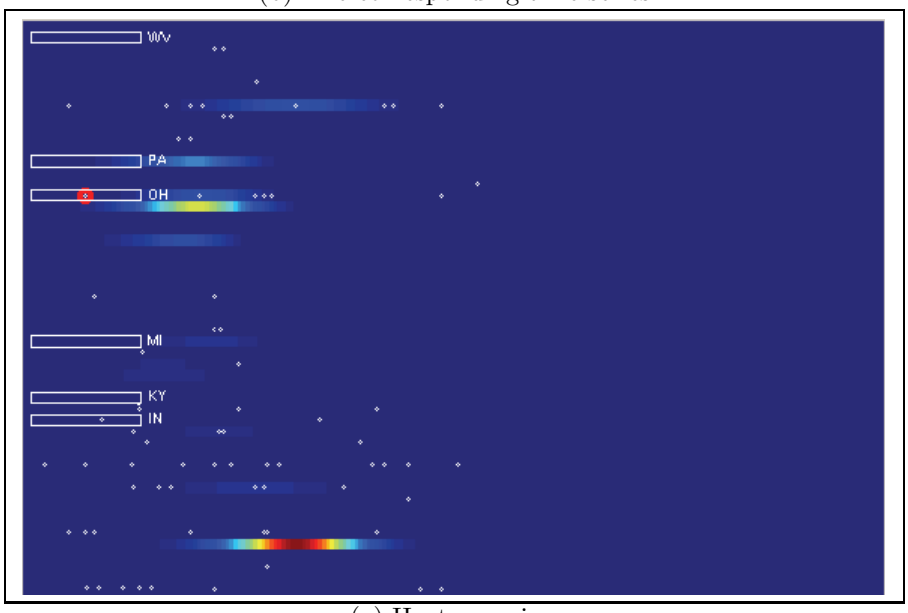
Please place your code in a directory titled "Code", the outputs to a directory called "Outputs", and your report in a directory called "Report"; zip or tar all off them together and submit it through the digital dropbox.

iteration	time	US-AK	US-AL	US-AR	US-AZ	US-CA	US-CO	US-CT	US-DC	US-DE	US-FL
1	1/1/2012 12:00	0	0	0	0	0	0	60.73668	0	0	0
2	1/3/2012 12:00	0	0	0	0	0	0	42.94187	0	0	0
3	1/4/2012 12:00	0.01201	0.047546	0.01743	0.090407	0.374936	0.098528	58.77457	0.020224	1.218536	0.285863
4	1/5/2012 12:00	0.013466	0.053334	0.019597	0.101359	0.421071	0.110305	75.78161	0.022414	2.326712	0.320572
5	1/6/2012 12:00	0.039996	0.162742	0.059873	0.297879	1.251812	0.325341	95.50366	0.211412	3.865487	0.9509
6	1/7/2012 12:00	0.05242	0.217835	0.080432	0.387193	1.645217	0.42332	118.7742	0.539795	5.94115	1.245282
7	1/8/2012 12:00	0.093136	0.394825	0.146143	0.681985	2.926991	0.747386	146.4766	1.125696	8.691269	2.210177
8	1/9/2012 12:00	0.123924	0.539149	0.200518	0.896928	3.905734	0.985088	179.4123	1.978989	12.2528	2.936908
9	1/10/2012 12:00	0.1884	0.832998	0.310841	1.352683	5.94622	1.48842	218.7079	3.257303	16.82901	4.460237
10	1/11/2012 12:00	0.24976	1.129475	0.423664	1.772906	7.902909	1.954983	265.527	5.009628	22.63206	5.904359
11	1/12/2012 12:00	0.353476	1.620949	0.610275	2.490014	11.20064	2.749909	321.4465	7.440166	29.9539	8.347058
12	1/13/2012 12:00	0.464983	2.173137	0.822204	3.24168	14.76656	3.586471	388.1729	10.65438	39.11295	10.96437
13	1/14/2012 12:00	0.632634	2.994778	1.137203	4.378379	20.1191	4.84996	467.935	14.92519	50.52897	14.90088
14	1/15/2012 12:00	0.82538	3.971995	1.514887	5.659622	26.29988	6.27802	563.2195	20.4412	64.67409	19.41405
15	1/16/2012 12:00	1.096039	5.339688	2.043401	7.463615	34.97282	8.286848	677.1898	27.57605	82.15283	25.75203
16	1/17/2012 12:00	1.419046	7.016697	2.695409	9.583676	45.35846	10.65174	813.4505	36.63998	103.6592	33.29948
17	1/18/2012 12:00	1.410498	8.966523	3.375072	12.0768	59.08338	13.35366	933.4336	35.52425	105.1526	43.0569
18	1/19/2012 12:00	1.246597	11.28319	4.153187	14.95383	75.75872	16.44539	1068.259	30.11702	100.1735	54.78521
19	1/20/2012 12:00	0.836845	14.1421	5.05853	18.5059	97.41766	20.1756	1210.514	17.97413	84.35114	69.86054
20	1/21/2012 12:00	0.169777	17.4474	6.02849	22.51139	123.82	24.25661	1357.567	1.430994	55.35926	87.94237
21	1/22/2012 12:00	0.017744	21.37153	7.085827	27.23977	157.1612	28.91068	1506.803	2.103989	11.63309	110.458

(a) Sample data



(b) The corresponding time series



(a) Heatmap view

Figure 1: A sample visualization showing a heatmap and a highlighted window: the state is “OH” and its 1-hop neighbors are “WV”, “PA”, “MI”, “KY”, and “IN”