Metastatic progression via biased random walks on a cancer network

P.K. Newton

Viterbi School of Engineering
&
Department of Mathematics
University of Southern California

Outline

Goals of the model:

- Build a computational 'platform' for Monte Carlo simulations of cancer progression
- Start by building this for the generic 'average' patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform 'tests'

Goals of the model:

- Build a computational 'platform' for Monte Carlo simulations of cancer progression
- Start by building this for the generic 'average' patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform 'tests'

Goals of the model:

- Build a computational 'platform' for Monte Carlo simulations of cancer progression
- Start by building this for the generic 'average' patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform 'tests'

Goals of the model:

- Build a computational ‘platform’ for Monte Carlo simulations of cancer progression
- Start by building this for the generic ‘average’ patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform ‘tests’

Goals of the model:

- Build a computational ‘platform’ for Monte Carlo simulations of cancer progression
- Start by building this for the generic ‘average’ patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform ‘tests’

Goals of the model:

- Build a computational 'platform' for Monte Carlo simulations of cancer progression
- Start by building this for the generic 'average' patient, focusing on lung cancer
- Then build the model for other types of cancer for comparison purposes
- Compare with individual patient histories
- Use model to quantify predictions
- Use model to perform 'tests'

DiSibio & French (2008)

3827 untreated patients (1914-1943)

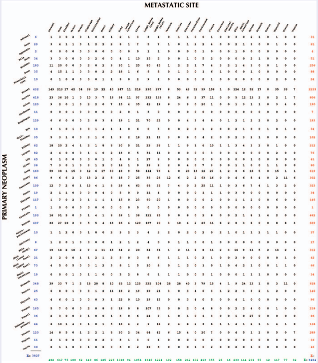


Figure 1. Data table for analyzed cases (for description, see "Results").

DiSibio & French (2008)

3827 untreated patients (1914-1943)

METASTATIC SITE

		Adrenal	Bone	Bladder	Brain	Breast	Diaphragm	Gallbladder	Heart	Kidney	Lung	Large intestine	Liver	Lymph Nodes (reg)	Lymph Nodes (dist)	Omentum	Ovaries	Pancreas	Pericardium	Peritoneum	Pleura	Prostate	Skeletal muscle	Skin	Small intestine	Spleen	Stomach	Testes	Thyroid	Uterus	Vagina	
Adrenal	6	1	3	0	2	0	1	0	0	1	2	1	3	3	4	0	1	1	0	0	1	0	0	3	0	2	1	0	1	0	0	31
Anus	29	3	4	1	1	0	1	2	2	2	8	1	7	5	7	1	0	1	2	2	4	0	0	2	1	3	0	0	1	0	0	61
Appendix	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	4
Bile duct	34	3	3	0	0	0	0	2	0	0	4	1	10	15	2	0	0	1	0	5	2	0	0	0	0	2	1	0	0	0	0	51
Bladder	183	11	20	0	0	0	2	0	2	9	30	1	25	80	45	1	2	2	1	7	4	3	2	1	4	3	0	0	1	0	0	256
Bone	35	4	15	1	1	0	3	0	2	2	18	1	6	8	8	0	1	3	0	1	6	1	0	6	0	1	0	0	0	0	0	88
Branchial cyst	10	0	1	0	0	0	0	0	1	1	3	0	2	9	4	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	24
Breast	432	149	213	17	42	54	36	19	22	40	247	11	218	230	277	8	53	49	52	59	158	1	0	124	12	52	17	0	35	33	7	2235
Cervix	418	23	36	10	1	0	10	3	7	19	94	11	97	232	133	8	24	6	2	37	11	0	0	5	12	13	2	0	2	1	7	806
Colon	123	9	2	0	1	0	2	2	0	7	15	4	35	42	19	6	3	9	0	20	1	0	0	1	3	1	1	0	3	4	0	190
Duodenum	11	0	0	0	1	0	0	0	0	0	2	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
Esophagus	129	6	6	0	0	0	2	0	3	4	19	1	21	70	22	0	0	4	3	4	8	0	1	2	1	2	2	0	2	0	0	183
Eye	10	3	1	0	1	0	0	1	4	1	4	0	6	0	3	0	0	2	0	1	2	0	0	2	1	0	0	1	0	1	0	34
Gall-bladder	35	5	3	1	0	0	3	1	0	1	9	2	18	21	13	3	0	8	0	4	2	0	0	2	3	2	1	0	0	0	0	102
Kidney	62	18	20	2	4	1	2	1	8	8	30	5	21	23	26	1	1	9	1	6	10	1	1	3	4	3	2	0	1	0	0	212

Key features of data set

- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the ‘long-time’ steady-state
- Patients are untreated males and females

Key features of data set

- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the 'long-time' steady-state
- Patients are untreated males and females

Key features of data set

- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the 'long-time' steady-state
- Patients are untreated males and females

Key features of data set

- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the ‘long-time’ steady-state
- Patients are untreated males and females

Key features of data set

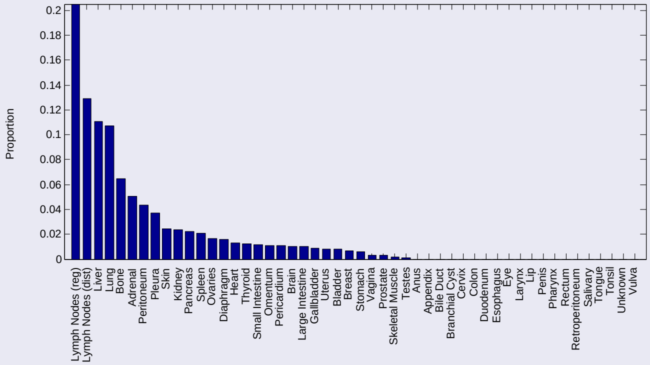
- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the ‘long-time’ steady-state
- Patients are untreated males and females

Key features of data set

- 50 combined primary or metastatic sites
- 39 primaries, 11 metastatic sites that are not primary sites
- Data set is large, but not comprehensive
- Each row gives ensemble metastatic distribution from a given primary
- These distributions represent the ‘long-time’ steady-state
- Patients are untreated males and females

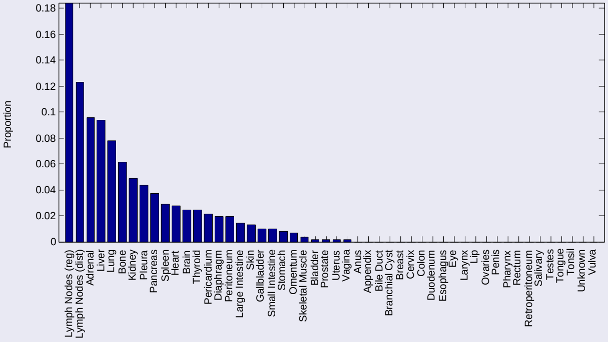
Metastatic distribution from all primaries

Overall metastatic distribution



Metastatic distribution from lung cancer primary

$\vec{v}_T$: The target 'statistical' steady-state distribution (27 sites)



Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the ‘nodes’ of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the 'nodes' of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

Defining the transition matrix

The Markov chain model

$$\vec{v}_{k+1} = \vec{v}_k A, \quad (k = 0, 1, 2, ...)$$

- 50 locations that are either primaries or metastatic sites
- A is a 50×50 transition matrix
- Rows sum to 1
- Entries are all primary and metastatic sites
- These will be the ‘nodes’ of our network model
- The nodes will be connected by directed edges
- Edge weightings will be determined by solving an optimization problem

The 50 nodes that make up our network

#	Name	#	Name
1	Adrenal	26	Omentum
2	Anus	27	Ovaries
3	Appendix	28	Pancreas
4	Bile Duct	29	Penis
5	Bladder	30	Pericardium
6	Bone	31	Peritoneum
7	Brain	32	Pharynx
8	Branchial Cyst	33	Pleura
9	Breast	34	Prostate
10	Cervix	35	Rectum
11	Colon	36	Retroperitoneum
12	Diaphragm	37	Salivary
13	Duodenum	38	Skeletal Muscle
14	Esophagus	39	Skin
15	Eye	40	Small Intestine
16	Gallbladder	41	Spleen
17	Heart	42	Stomach
18	Kidney	43	Testes
19	Large Intestine	44	Thyroid
20	Larynx	45	Tongue
21	Lip	46	Tonsil
22	Liver	47	Unknown
23	Lung	48	Uterus
24	Lymph Nodes (reg)	49	Vagina
25	Lymph Nodes (dist)	50	Vulva

State-vector representation

$\vec{v}_0$: The initial state-vector

- Represents the distribution of primary tumors (and our level of certainty)
- $\vec{v}_0 = (1, 0, 0, 0, ...)$: Primary tumor located in Adrenal gland
- $\vec{v}_0 = (0, 0, 0, 0, ..., 1, 0, 0, 0, ...)$: Primary tumor located in Lung
- $\vec{v}_0 = (1/50, 1/50, 1/50, ...)$: Complete lack of information on location of primary tumor
- $\vec{v}_0 = (1/2, 0, 0, 0, ..., 1/2, 0, 0, 0, ...)$: Primary tumor located in Adrenal and/or Lung

State-vector representation

$\vec{v}_0$: The initial state-vector

- Represents the distribution of primary tumors (and our level of certainty)
- $\vec{v}_0 = (1, 0, 0, 0, ...)$: Primary tumor located in Adrenal gland
- $\vec{v}_0 = (0, 0, 0, 0, ..., 1, 0, 0, 0, ...)$: Primary tumor located in Lung
- $\vec{v}_0 = (1/50, 1/50, 1/50, ...)$: Complete lack of information on location of primary tumor
- $\vec{v}_0 = (1/2, 0, 0, 0, ..., 1/2, 0, 0, 0, ...)$: Primary tumor located in Adrenal and/or Lung

State-vector representation

$\vec{v}_0$: The initial state-vector

- Represents the distribution of primary tumors (and our level of certainty)
- $\vec{v}_0 = (1, 0, 0, 0, ...)$: Primary tumor located in Adrenal gland
- $\vec{v}_0 = (0, 0, 0, 0, ..., 1, 0, 0, 0, ...)$: Primary tumor located in Lung
- $\vec{v}_0 = (1/50, 1/50, 1/50, ...)$: Complete lack of information on location of primary tumor
- $\vec{v}_0 = (1/2, 0, 0, 0, ..., 1/2, 0, 0, 0, ...)$: Primary tumor located in Adrenal and/or Lung

State-vector representation

$\vec{v}_0$: The initial state-vector

- Represents the distribution of primary tumors (and our level of certainty)
- $\vec{v}_0 = (1, 0, 0, 0, ...)$: Primary tumor located in Adrenal gland
- $\vec{v}_0 = (0, 0, 0, 0, ..., 1, 0, 0, 0, ...)$: Primary tumor located in Lung
- $\vec{v}_0 = (1/50, 1/50, 1/50, ...)$: Complete lack of information on location of primary tumor
- $\vec{v}_0 = (1/2, 0, 0, 0, ..., 1/2, 0, 0, 0, ...)$: Primary tumor located in Adrenal and/or Lung

State-vector representation

$\vec{v}_0$: The initial state-vector

- Represents the distribution of primary tumors (and our level of certainty)
- $\vec{v}_0 = (1, 0, 0, 0, ...)$: Primary tumor located in Adrenal gland
- $\vec{v}_0 = (0, 0, 0, 0, ..., 1, 0, 0, 0, ...)$: Primary tumor located in Lung
- $\vec{v}_0 = (1/50, 1/50, 1/50, ...)$: Complete lack of information on location of primary tumor
- $\vec{v}_0 = (1/2, 0, 0, 0, ..., 1/2, 0, 0, 0, ...)$: Primary tumor located in Adrenal and/or Lung

Metastatic progression

State-vector dynamics

$$
\begin{aligned}
\vec{v}_1 &= \vec{v}_0 A \\
\vec{v}_2 &= \vec{v}_1 A = \vec{v}_0 A^2 \\
\vec{v}_3 &= \vec{v}_2 A = \vec{v}_0 A^3 \\
&\vdots \\
\vec{v}_{k+1} &= \vec{v}_k A = \vec{v}_0 A^{k+1}
\end{aligned}
$$

The steady-state ‘statistical’ distribution

$\vec{v}_\infty$: $k \to \infty$

$$\vec{v}_\infty = \lim_{k\to\infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

The steady-state 'statistical' distribution

$\vec{v}_\infty$: $k \to \infty$

$$\vec{v}_\infty = \lim_{k\to\infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

The steady-state ‘statistical’ distribution

$\vec{v}_\infty$: $k \to \infty$

$$\vec{v}_\infty = \lim_{k \to \infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

The steady-state ‘statistical’ distribution

$\vec{v}_\infty$: $k \to \infty$

$$\vec{v}_\infty = \lim_{k\to\infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

The steady-state ‘statistical’ distribution

$\vec{v}_\infty$: $k \rightarrow \infty$

$$\vec{v}_\infty = \lim_{k \rightarrow \infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

The steady-state 'statistical' distribution

$\vec{v}_\infty$: $k \to \infty$

$$\vec{v}_\infty = \lim_{k\to\infty} \vec{v}_0 A^k$$

$$\vec{v}_\infty = \vec{v}_\infty A$$

$$\vec{v}_\infty (A - I) = 0$$

- $\vec{v}_\infty$ is an eigenvector of A corresponding to eigenvalue 1
- Since A is stochastic, must have at least one steady-state
- Find entries of A so that $\vec{v}_\infty \equiv \vec{v}_T$

Constrained linear optimization problem

Find the entries a_{ij} of the transition matrix A, subject to:

$$\vec{v}_T(A - I) = 0$$

- Constraints: $0 \leq a_{ij} \leq 1$; $\sum_{j=1}^{50} a_{ij} = 1$.

Constrained linear optimization problem

Find the entries a_{ij} of the transition matrix A, subject to:

$$\vec{v}_T(A - I) = 0$$

- Constraints: $0 \leq a_{ij} \leq 1$; $\sum_{j=1}^{50} a_{ij} = 1$.

'Training' the matrix

A_j is a 50×50 transition matrix

$$
\begin{aligned}
\vec{v}_T(A_j - I) &\neq 0 \\
\vec{v}_\infty^{(j)}(A_j - I) &= 0
\end{aligned}
$$

$\vec{r}_j$: The residual vector

$$
\vec{v}_T(A_j - I) = (\vec{v}_T - \vec{v}_\infty^{(j)})(A_j - I) \equiv \vec{r}_j
$$

'Training' the matrix

A_j is a 50×50 transition matrix

$$\begin{aligned} \vec{v}_T(A_j - I) &\neq 0 \\ \vec{v}_\infty^{(j)}(A_j - I) &= 0 \end{aligned}$$

$\vec{r}_j$: The residual vector

$$\vec{v}_T(A_j - I) = (\vec{v}_T - \vec{v}_\infty^{(j)})(A_j - I) \equiv \vec{r}_j$$

‘Training’ the matrix

Two-step process

- Step 1: Approximate (A_0, $j = 0$)
- Step 2: Iterate (A_j, $j = 1, 2, ...$)

- Goal: Drive the residual norm $\|\vec{r}_j\|^2$ to zero as $j \to \infty$.
- $\|(\vec{v}_T - \vec{v}_\infty^{(j)})\|^2 \to 0$

'Training' the matrix

Two-step process

- Step 1: Approximate ($A_0, j = 0$)
- Step 2: Iterate ($A_j, j = 1, 2, ...$)

- Goal: Drive the residual norm $\|\vec{r}_j\|^2$ to zero as $j \to \infty$.
- $\|(\vec{v}_T - \vec{v}_\infty^{(j)})\|^2 \to 0$

'Training' the matrix

Two-step process

- Step 1: Approximate (A_0, $j = 0$)
- Step 2: Iterate (A_j, $j = 1, 2, ...$)

- Goal: Drive the residual norm $\|\vec{r}_j\|^2$ to zero as $j \to \infty$.
- $\|(\vec{v}_T - \vec{v}_\infty^{(j)})\|^2 \to 0$

Finding a 'good' A_0

Our first approximation

- For each primary in data set, use raw numbers to fill out rows
- Normalize each of those rows by dividing by sum
- Fill out entries for the remaining empty rows with 'unbiased' values $a_{ij} = 1/50$
- A_0 constructed this way is a stochastic transition matrix
- But $\vec{v}_T(A_0 - I) \neq 0$

Finding a ‘good’ A_0

Our first approximation

- For each primary in data set, use raw numbers to fill out rows
- Normalize each of those rows by dividing by sum
- Fill out entries for the remaining empty rows with ‘unbiased’ values $a_{ij} = 1/50$
- A_0 constructed this way is a stochastic transition matrix
- But $\vec{v}_T(A_0 - I) \neq 0$

Finding a ‘good’ A_0

Our first approximation

- For each primary in data set, use raw numbers to fill out rows
- Normalize each of those rows by dividing by sum
- Fill out entries for the remaining empty rows with ‘unbiased’ values $a_{ij} = 1/50$
- A_0 constructed this way is a stochastic transition matrix
- But $\vec{v}_T(A_0 - I) \neq 0$

Finding a 'good' A_0

Our first approximation

- For each primary in data set, use raw numbers to fill out rows
- Normalize each of those rows by dividing by sum
- Fill out entries for the remaining empty rows with 'unbiased' values $a_{ij} = 1/50$
- A_0 constructed this way is a stochastic transition matrix
- But $\vec{v}_T(A_0 - I) \neq 0$

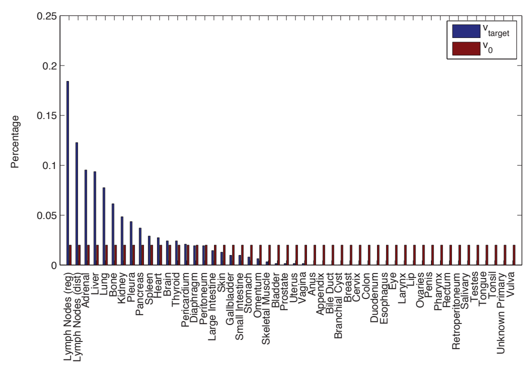
Finding a 'good' A_0

Our first approximation

- For each primary in data set, use raw numbers to fill out rows
- Normalize each of those rows by dividing by sum
- Fill out entries for the remaining empty rows with 'unbiased' values $a_{ij} = 1/50$
- A_0 constructed this way is a stochastic transition matrix
- But $\vec{v}_T(A_0 - I) \neq 0$

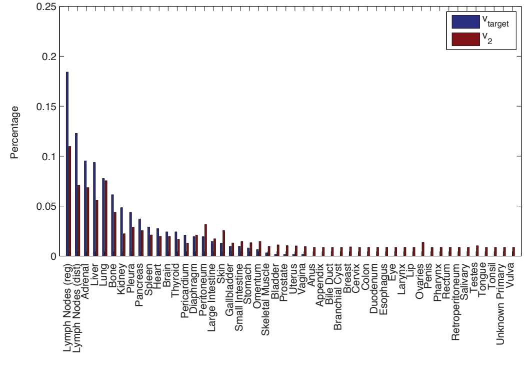
Dynamics using A_0

$k = 0$



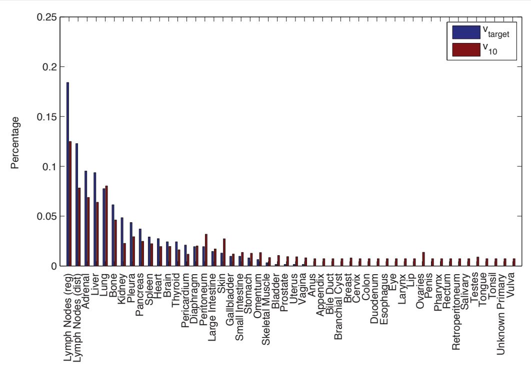
Dynamics using A_0

$k = 2$



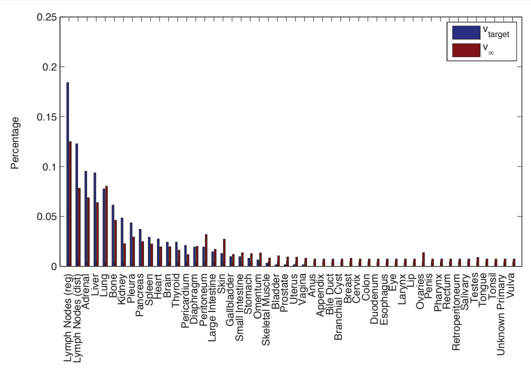
Dynamics using A_0

$k = 10$



Dynamics using A_0

$k = \infty$: Does not converge to the correct steady-state



Iterating A_j, $j = 0, 1, 2, \ldots$

The iteration scheme

1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $\|\vec{r}_j\|^2$. The new matrix is A_{j+1}.
6. Stop if $\|\vec{r}_{j+1}\|^2 < \epsilon$, otherwise go to step 2 and repeat.

Iterating A_j, $j = 0, 1, 2, ...$

The iteration scheme

1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $\|\vec{r}_j\|^2$. The new matrix is A_{j+1}.
6. Stop if $\|\vec{r}_{j+1}\|^2 < \epsilon$, otherwise go to step 2 and repeat.

Iterating A_j, $j = 0, 1, 2, \ldots$

The iteration scheme

1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $\|\vec{r}_j\|^2$. The new matrix is A_{j+1}.
6. Stop if $\|\vec{r}_{j+1}\|^2 < \epsilon$, otherwise go to step 2 and repeat.

Iterating A_j, $j = 0, 1, 2, \ldots$

The iteration scheme

1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $||\vec{r}_j||^2$. The new matrix is A_{j+1}.
6. Stop if $||\vec{r}_{j+1}||^2 < \epsilon$, otherwise go to step 2 and repeat.

Iterating A_j, $j = 0, 1, 2, \ldots$

The iteration scheme

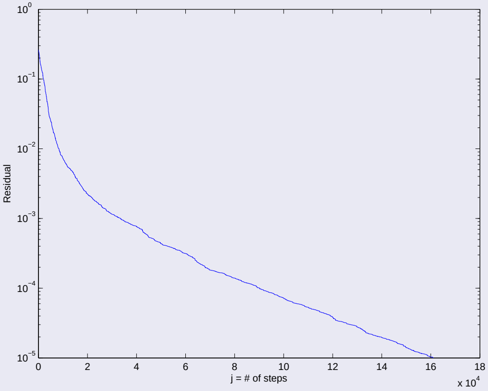
1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $\|\vec{r}_j\|^2$. The new matrix is A_{j+1}.
6. Stop if $\|\vec{r}_{j+1}\|^2 < \epsilon$, otherwise go to step 2 and repeat.

Iterating A_j, $j = 0, 1, 2, \ldots$

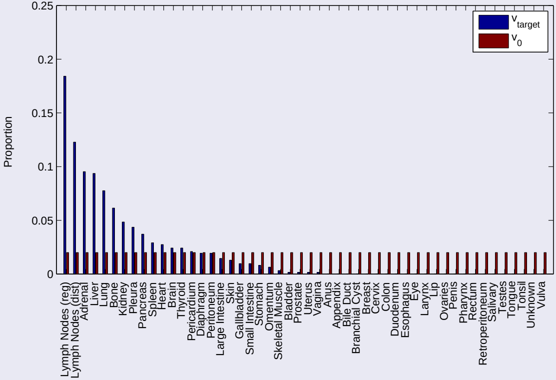
The iteration scheme

1. Calculate the residual $\vec{r}_j$ at step j
2. Pick column of A_j corresponding to position of max entry of $\vec{r}_j$
3. Pick column of A_j corresponding to position of min entry of $\vec{r}_j$
4. Pick a row of A_j at random
5. Decrease entry of A_j in step 2 by δ, increase entry of A_j in step 3 by δ, where δ scales with $\|\vec{r}_j\|^2$. The new matrix is A_{j+1}.
6. Stop if $\|\vec{r}_{j+1}\|^2 < \epsilon$, otherwise go to step 2 and repeat.

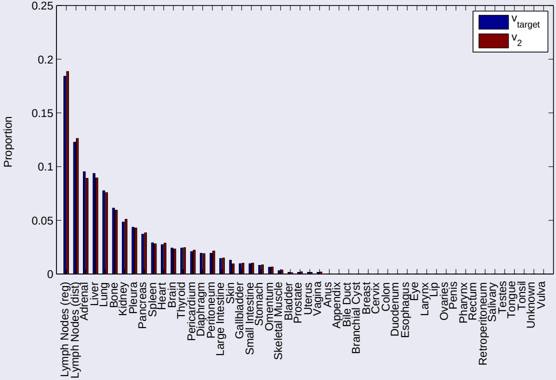
Convergence to the fully ‘trained’ matrix



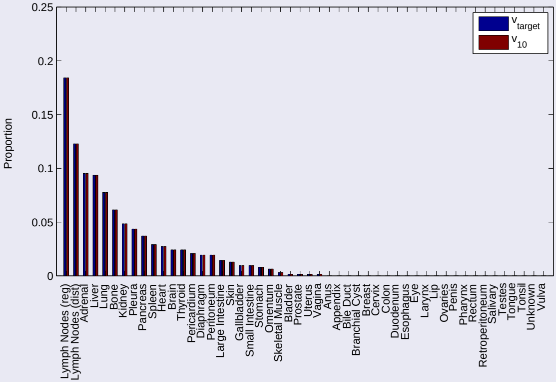
$k = 0$: State-vector dynamics using trained lung cancer matrix



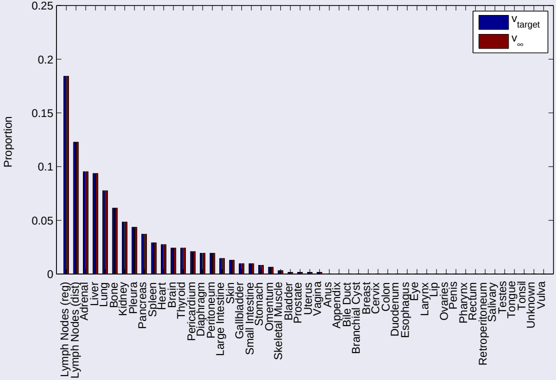
$k = 2$: State-vector dynamics using trained lung cancer matrix



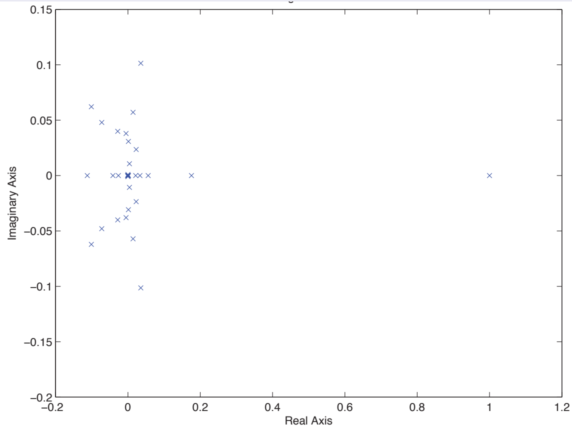
$k = 10$: State-vector dynamics using trained lung cancer matrix



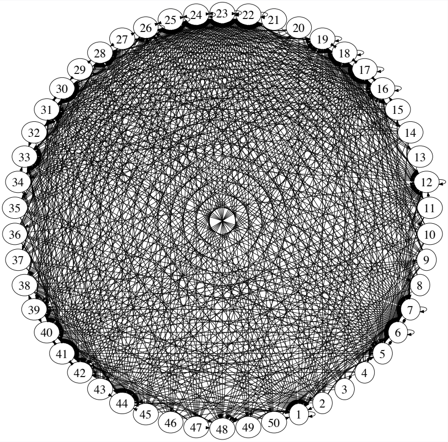
$k = \infty$: State-vector dynamics using trained lung cancer matrix



Eigenvalue distribution of lung cancer matrix

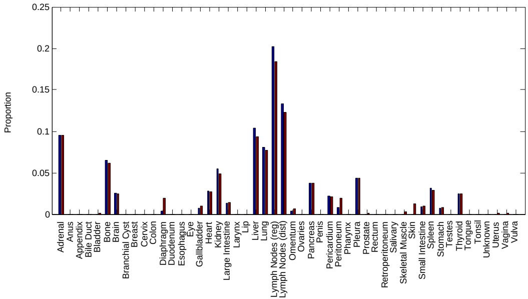


The lung cancer network



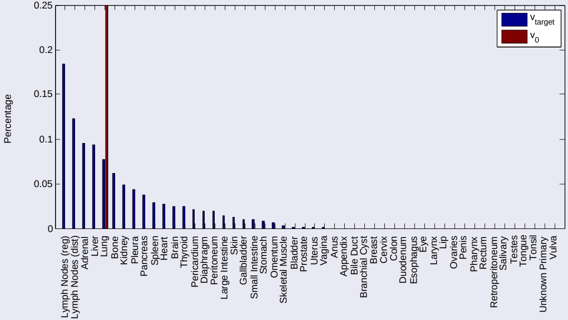
Summary of network structure

- Total of 913 edges
- Lung node has 21 outgoing edges
- Lung node has 49 incoming edges



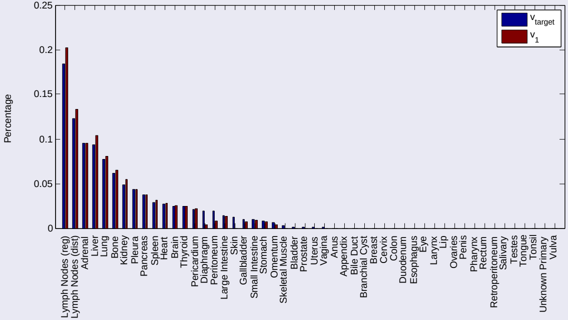
Importance of secondary connections

Do the first outgoing edges from the lung tell the whole story?



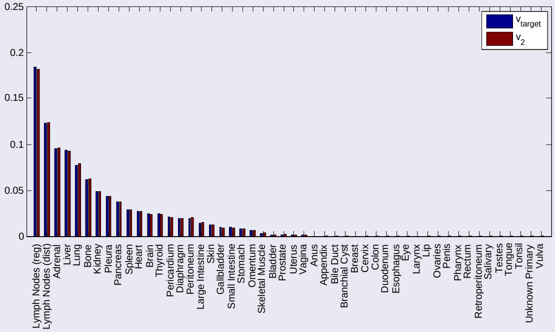
Importance of secondary connections

$k = 1$: 21 ‘First order’ connections



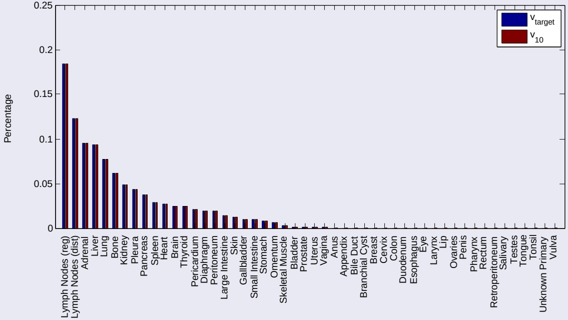
Importance of secondary connections

$k = 2$: 6 ‘Second order connections’ - ‘mets from mets’



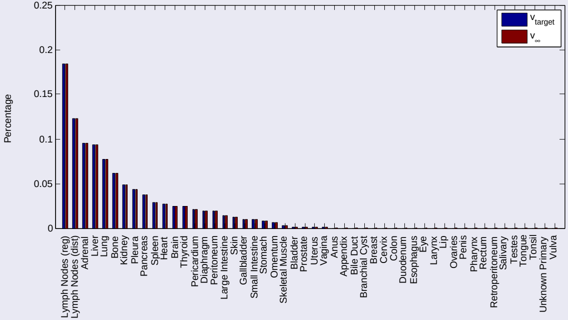
Importance of secondary connections

$k = 10$



Importance of secondary connections

$k = \infty$

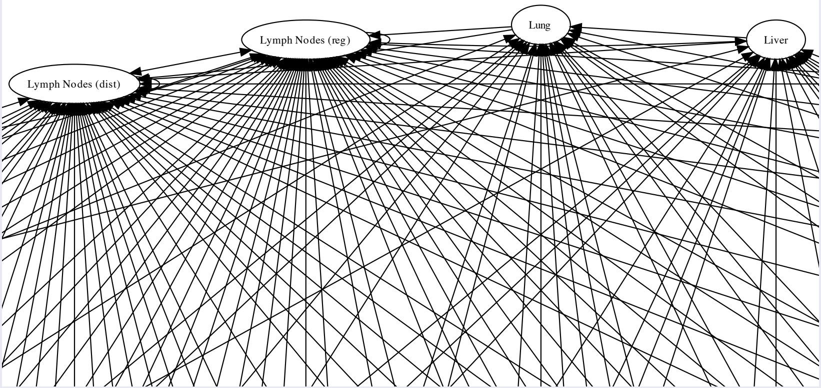


‘Second order’ cancers (from primary lung)

Mets from mets

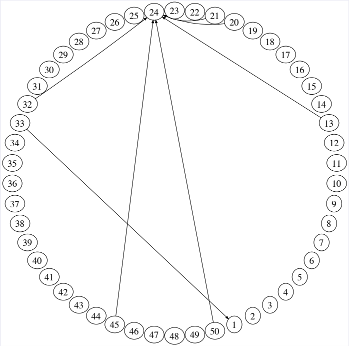
- bladder
- prostate
- skeletal muscle
- skin
- uterus
- vagina

The most heavily weighted incoming connections



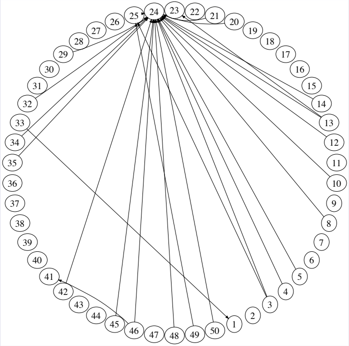
Cut-off: 0.4

Strongest connections all go to the (24) Lymph nodes (reg) and (33) Pleura to (1) Adrenal.



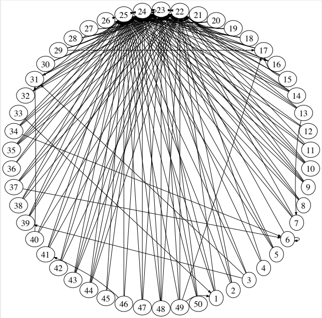
Cut-off: 0.25

Strongest connections to (24) Lymph nodes (reg), (25) Lymph nodes (dist), (23) Lung, and (46) Tonsil to (41) Spleen.



Cut-off: 0.1

Connections to (24) Lymph nodes (reg), (25) Lymph nodes (dist), (23) Lung, (22) Liver, new connections to (1) Adrenal, (6) Bone, (7) Brain, and (17) Heart.



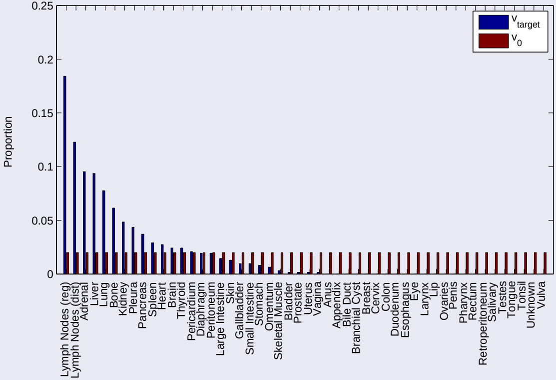
Question: Would a 'pure diffusion' process work?

- Suppose we replace the heterogeneous edge weightings with 'unbiased' weighting where edge weights are distributed equally across all outgoing edges at each node.

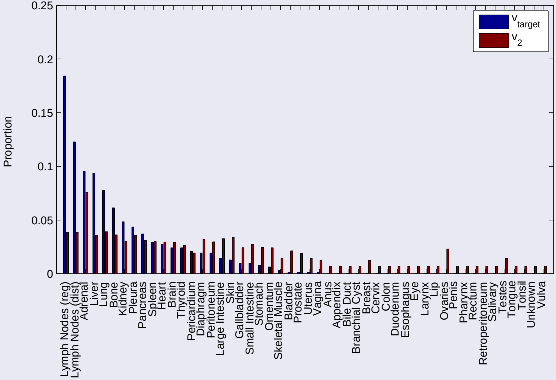
Question: Would a ‘pure diffusion’ process work?

- Suppose we replace the heterogeneous edge weightings with ‘unbiased’ weighting where edge weights are distributed equally across all outgoing edges at each node.

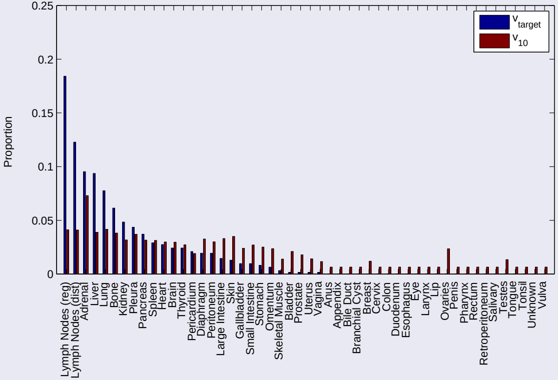
$k = 0$: State-vector dynamics using unbiased edge weightings



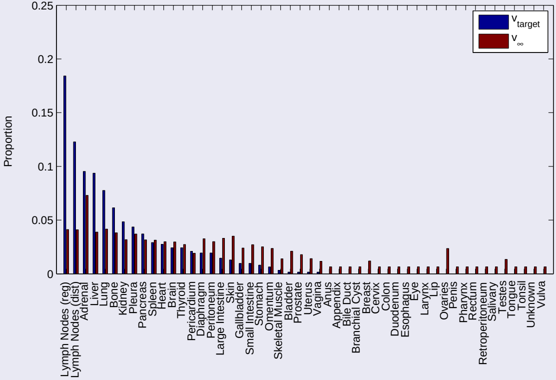
$k = 2$: State-vector dynamics using unbiased edge weightings



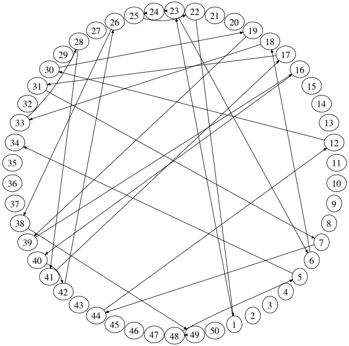
$k = 10$: State-vector dynamics using unbiased edge weightings



$k = \infty$: Does not converge to correct steady-state



One ‘Monte Carlo’ trajectory from the lung



- How many steps (on average) does it take to go from lung $\rightarrow$ node i ?

Mean first-passage time matrix

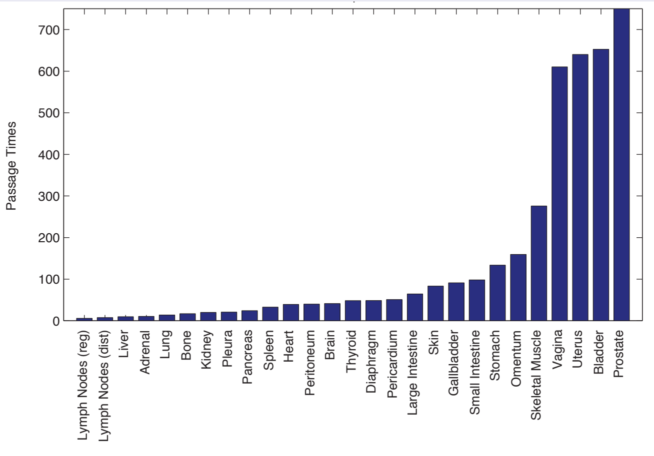
$$
\begin{aligned}
Z &= (I - P + W)^{-1} \\
I &: \quad n \times n \quad \text{identity} \quad \text{matrix} \\
P &: \quad n \times n \quad \text{transition} \quad \text{matrix} \\
W &: \quad \text{rows} \quad \text{are} \quad \text{steady-state}
\end{aligned}
$$

$$
m_{ij} = \frac{z_{jj} - z_{ij}}{w_j}
$$

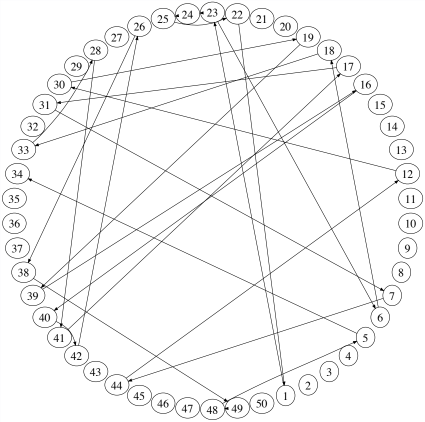
Z matrix also gives variances

Mean First Passage Time from Lung		Variance
Lymph Nodes (reg)	5.3295	4.86733
Lymph Nodes (dist)	7.8069	7.38658
Liver	9.7405	9.21283
Adrenal	9.9006	9.38006
Lung	12.8793	12.5152
Bone	18.3202	18.2009
Kidney	20.1983	19.9714
Pleura	21.9595	21.368
Pancreas	26.0553	25.4704
Spleen	34.7067	34.1273
Heart	36.6631	35.8982
Thyroid	40.4995	39.5196
Brain	40.9396	41.1525
Pericardium	48.652	46.7418
Peritoneum	51.0337	50.0885
Diaphragm	52.1855	50.7323
Large Intestine	68.9146	68.2363
Skin	79.334	77.3178
Gallbladder	104.151	101.483
Small Intestine	104.491	102.993
Stomach	122.915	122.968
Omentum	156.996	155.52
Skeletal Muscle	308.253	308.538
Uterus	604.221	600.579
Bladder	614.423	622.782
Prostate	619.438	628.354
Vagina	629.237	642.475

Mean first-passage times



The mean first-passage time trajectory



Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

Mean first-passage time ordering

Lung → Lymph nodes (reg) (1 time unit)
Lung → Lymph nodes (dist) (1.46 time units)
Lung → Liver (1.83 time units)
Lung → Adrenal (1.86 time units)
Lung → Lung (2.42 time units)
Lung → Bone (3.44 time units)
Lung → Kidney (3.79 time units)
Lung → Pleura (4.12 time units)
Lung → Pancreas (4.89 time units)

Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

Mean first-passage time ordering

Lung → Lymph nodes (reg) (1 time unit)
Lung → Lymph nodes (dist) (1.46 time units)
Lung → Liver (1.83 time units)
Lung → Adrenal (1.86 time units)
Lung → Lung (2.42 time units)
Lung → Bone (3.44 time units)
Lung → Kidney (3.79 time units)
Lung → Pleura (4.12 time units)
Lung → Pancreas (4.89 time units)

Mean first-passage time ordering

Lung → Lymph nodes (reg) (1 time unit)
Lung → Lymph nodes (dist) (1.46 time units)
Lung → Liver (1.83 time units)
Lung → Adrenal (1.86 time units)
Lung → Lung (2.42 time units)
Lung → Bone (3.44 time units)
Lung → Kidney (3.79 time units)
Lung → Pleura (4.12 time units)
Lung → Pancreas (4.89 time units)

Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

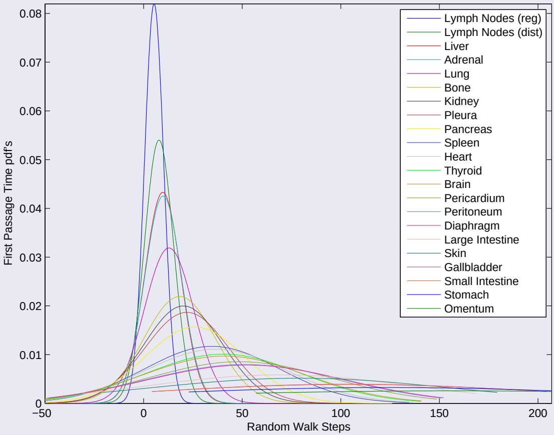
Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

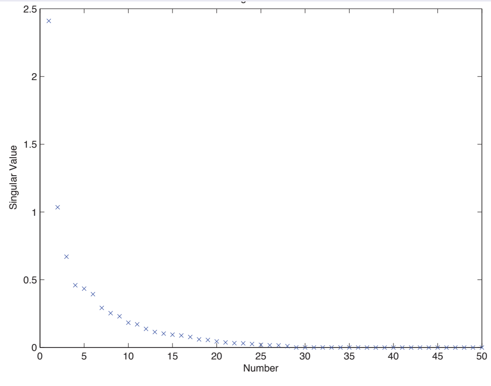
Mean first-passage time ordering

Lung $\rightarrow$ Lymph nodes (reg) (1 time unit)
Lung $\rightarrow$ Lymph nodes (dist) (1.46 time units)
Lung $\rightarrow$ Liver (1.83 time units)
Lung $\rightarrow$ Adrenal (1.86 time units)
Lung $\rightarrow$ Lung (2.42 time units)
Lung $\rightarrow$ Bone (3.44 time units)
Lung $\rightarrow$ Kidney (3.79 time units)
Lung $\rightarrow$ Pleura (4.12 time units)
Lung $\rightarrow$ Pancreas (4.89 time units)

Mean first-passage times are Gaussian distributed rv's



Singular value distribution of lung cancer matrix



- Eigenvalues of the 'covariance' matrix $A^T A$

Conclusions

Main points

- Metastatic progression can be thought of as a ‘biased’ random walk process on a network of potential metastatic sites.
- Pure diffusion process (unbiased random walk) is not a good model.
- Model identifies 21 ‘first-order’ sites, and 6 ‘second-order’ sites (‘mets from mets’)
- Model flags 9 sites (out of the 50 total number) with very short mean first-passage times.

Conclusions

Main points

- Metastatic progression can be thought of as a ‘biased’ random walk process on a network of potential metastatic sites.
- Pure diffusion process (unbiased random walk) is not a good model.
- Model identifies 21 ‘first-order’ sites, and 6 ‘second-order’ sites (‘mets from mets’)
- Model flags 9 sites (out of the 50 total number) with very short mean first-passage times.

Conclusions

Main points

- Metastatic progression can be thought of as a ‘biased’ random walk process on a network of potential metastatic sites.
- Pure diffusion process (unbiased random walk) is not a good model.
- Model identifies 21 ‘first-order’ sites, and 6 ‘second-order’ sites (‘mets from mets’)
- Model flags 9 sites (out of the 50 total number) with very short mean first-passage times.

Conclusions

Main points

- Metastatic progression can be thought of as a 'biased' random walk process on a network of potential metastatic sites.
- Pure diffusion process (unbiased random walk) is not a good model.
- Model identifies 21 'first-order' sites, and 6 'second-order' sites ('mets from mets')
- Model flags 9 sites (out of the 50 total number) with very short mean first-passage times.

Conclusions

Secondary points

- Complex web of connections are important, not just outgoing ones from lung.
- Model supports the idea that metastatic development is the result of a complex and intricate pattern of cross-talk and communication among a large collection of potential nodes.
- Next: Comparisons of different cancer networks (lung, liver, breast, colon, prostate, ovarian)
- Individual patient histories and Bayesian updating

Conclusions

Secondary points

- Complex web of connections are important, not just outgoing ones from lung.
- Model supports the idea that metastatic development is the result of a complex and intricate pattern of cross-talk and communication among a large collection of potential nodes.
- Next: Comparisons of different cancer networks (lung, liver, breast, colon, prostate, ovarian)
- Individual patient histories and Bayesian updating

Conclusions

Secondary points

- Complex web of connections are important, not just outgoing ones from lung.
- Model supports the idea that metastatic development is the result of a complex and intricate pattern of cross-talk and communication among a large collection of potential nodes.
- Next: Comparisons of different cancer networks (lung, liver, breast, colon, prostate, ovarian)
- Individual patient histories and Bayesian updating

Conclusions

Secondary points

- Complex web of connections are important, not just outgoing ones from lung.
- Model supports the idea that metastatic development is the result of a complex and intricate pattern of cross-talk and communication among a large collection of potential nodes.
- Next: Comparisons of different cancer networks (lung, liver, breast, colon, prostate, ovarian)
- Individual patient histories and Bayesian updating

References

- P.K. Newton, J. Mason, P. Kuhn, K. Bethel, J. Nieva, L. Bazhenova, *Metastatic progression via biased random walks on a cancer network*, USC preprint (2011).
- G. DiSibio, S.W. French [2008], Metastatic patterns of cancers: Results from a large autopsy study, *Arch Pathol Lab Med*, **Vol. 132**, June.
- KI Goh, ME Cusick, D Valle, B Childs, M Vidal, AL Barabasi [2007], The human disease network, *Proc. Nat'l Acad. Sci.* **104**, 8685-8690.
- J. Balthrop, S. Forrest, M.E.J. Newman [2004], Technological networks and the spread of computer viruses, *Science*, **304**, 5670, 527–529.
- NCI/NIH PS-OC: 'The Physics & Mathematics of Cancer Metastasis', 2010-2015.

References

- P.K. Newton, J. Mason, P. Kuhn, K. Bethel, J. Nieva, L. Bazhenova, *Metastatic progression via biased random walks on a cancer network*, USC preprint (2011).
- G. DiSibio, S.W. French [2008], Metastatic patterns of cancers: Results from a large autopsy study, *Arch Pathol Lab Med*, **Vol. 132**, June.
- KI Goh, ME Cusick, D Valle, B Childs, M Vidal, AL Barabasi [2007], The human disease network, *Proc. Nat'l Acad. Sci.* **104**, 8685-8690.
- J. Balthrop, S. Forrest, M.E.J. Newman [2004], Technological networks and the spread of computer viruses, *Science*, **304**, 5670, 527–529.
- NCI/NIH PS-OC: 'The Physics & Mathematics of Cancer Metastasis', 2010-2015.

References

- P.K. Newton, J. Mason, P. Kuhn, K. Bethel, J. Nieva, L. Bazhenova, *Metastatic progression via biased random walks on a cancer network*, USC preprint (2011).
- G. DiSibio, S.W. French [2008], Metastatic patterns of cancers: Results from a large autopsy study, *Arch Pathol Lab Med*, **Vol. 132**, June.
- KI Goh, ME Cusick, D Valle, B Childs, M Vidal, AL Barabasi [2007], The human disease network, *Proc. Nat'l Acad. Sci.* **104**, 8685-8690.
- J. Balthrop, S. Forrest, M.E.J. Newman [2004], Technological networks and the spread of computer viruses, *Science*, **304**, 5670, 527–529.
- NCI/NIH PS-OC: 'The Physics & Mathematics of Cancer Metastasis', 2010-2015.

References

- P.K. Newton, J. Mason, P. Kuhn, K. Bethel, J. Nieva, L. Bazhenova, *Metastatic progression via biased random walks on a cancer network*, USC preprint (2011).
- G. DiSibio, S.W. French [2008], Metastatic patterns of cancers: Results from a large autopsy study, *Arch Pathol Lab Med*, **Vol. 132**, June.
- KI Goh, ME Cusick, D Valle, B Childs, M Vidal, AL Barabasi [2007], The human disease network, *Proc. Nat'l Acad. Sci.* **104**, 8685-8690.
- J. Balthrop, S. Forrest, M.E.J. Newman [2004], Technological networks and the spread of computer viruses, *Science*, **304**, 5670, 527–529.
- NCI/NIH PS-OC: 'The Physics & Mathematics of Cancer Metastasis', 2010-2015.