

In [20]:

```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

Let's download some data from here: (old at the moment...who cares)

<http://blog.smellthedata.com/2012/03/data-for-2012.html> (<http://blog.smellthedata.com/2012/03/data-for-2012.html>)

<http://blog.smellthedata.com/2011/03/selection-sunday-today.html>
(<http://blog.smellthedata.com/2011/03/selection-sunday-today.html>)

Open up the `teams.tsv` file to get a list of NCAA basketball teams and their abbreviations:

In [21]:

```
teamfile = open('teams.tsv')
teams={}
teamsID={}
teamsName={}
k=0
for line in teamfile:
    (teamlabel,teamname)=line.split(None,1)
    teams[teamlabel]=teamname.strip()
    teamsID[teamlabel]=k
    teamsName[k]=teamname.strip()
    k+=1
teamfile.close()
nteam = len(teams)
```

We're going to look at maxgames of basketball scores

We will construct maxgames edges called `winlossIJ` and a weight for these edges called `winlossData`

In [22]:

```
maxgames = 100000
winlossIJ = np.zeros((maxgames,2),dtype=int) # winner loser
winlossData = np.zeros((maxgames,)) # weight
```

Now read the basketball scores.

For each line read as

- game date
- home team
- away team
- home score
- away score
- winner

We have to decide. How "much" of a win is something? We'll mark the winner as a 1 and the loser as a 0. If the score is close (less than `breakscore`) then give the winner a weight of 0.5 and the loser a weight of 0.4. Otherwise weight the score closer to 1.0.

In [23]:

```
k=0
ngames=0
breakscore = 6
gamesfile = open('GameResults_20110311.tsv')
for line in gamesfile:
    (gamedate,hteam,ateam,hscore,ascore,winner)=line.split()
    if hteam=='UNK' or ateam=='UNK':
        continue
    hscore=int(hscore)
    ascore=int(ascore)
    # winner = int(winner)
    gamedate=int(gamedate)
    if gamedate>=20111107 and gamedate<=20120311:
        if winner==1:
            wteam = hteam
            lteam = ateam
        else:
            lteam = hteam
            wteam = ateam
        scorediff = abs(hscore-ascore)
        if scorediff<breakscore:
            wweight = 0.5    # winner weight
            lweight = 0.4    # loser weight
            winlossIJ[k,:]=[teamsID[lteam],teamsID[wteam]]
            winlossData[k]=wweight
            k+=1
            winlossIJ[k,:]=[teamsID[wteam],teamsID[lteam]]
            winlossData[k]=lweight
        else:
            wweight = 1/(1+np.exp(-(scorediff-breakscore))) # logistic function
            lweight = 1-wweight
            winlossIJ[k,:]=[teamsID[lteam],teamsID[wteam]]
            winlossData[k]=wweight
        ngames+=1
        k+=1
    else:
        break
gamesfile.close()
print('parse %d games' % k)
```

parse 6606 games

Now we need to assemble the list of edges and weights as an $n \times n$ matrix.

In [32]:

```
winlossIJ=np.array(winlossIJ[:k,:])
winlossData=np.array(winlossData[:k])

import scipy.sparse as sparse
winlossmatrix=sparse.coo_matrix((winlossData,(winlossIJ[:,0],winlossIJ[:,1])))
```

In [33]:

```
winlossmatrix
```

Out[33]:

```
<346x346 sparse matrix of type '<class 'numpy.float64'>'
    with 6606 stored elements in COOrdinate format>
```

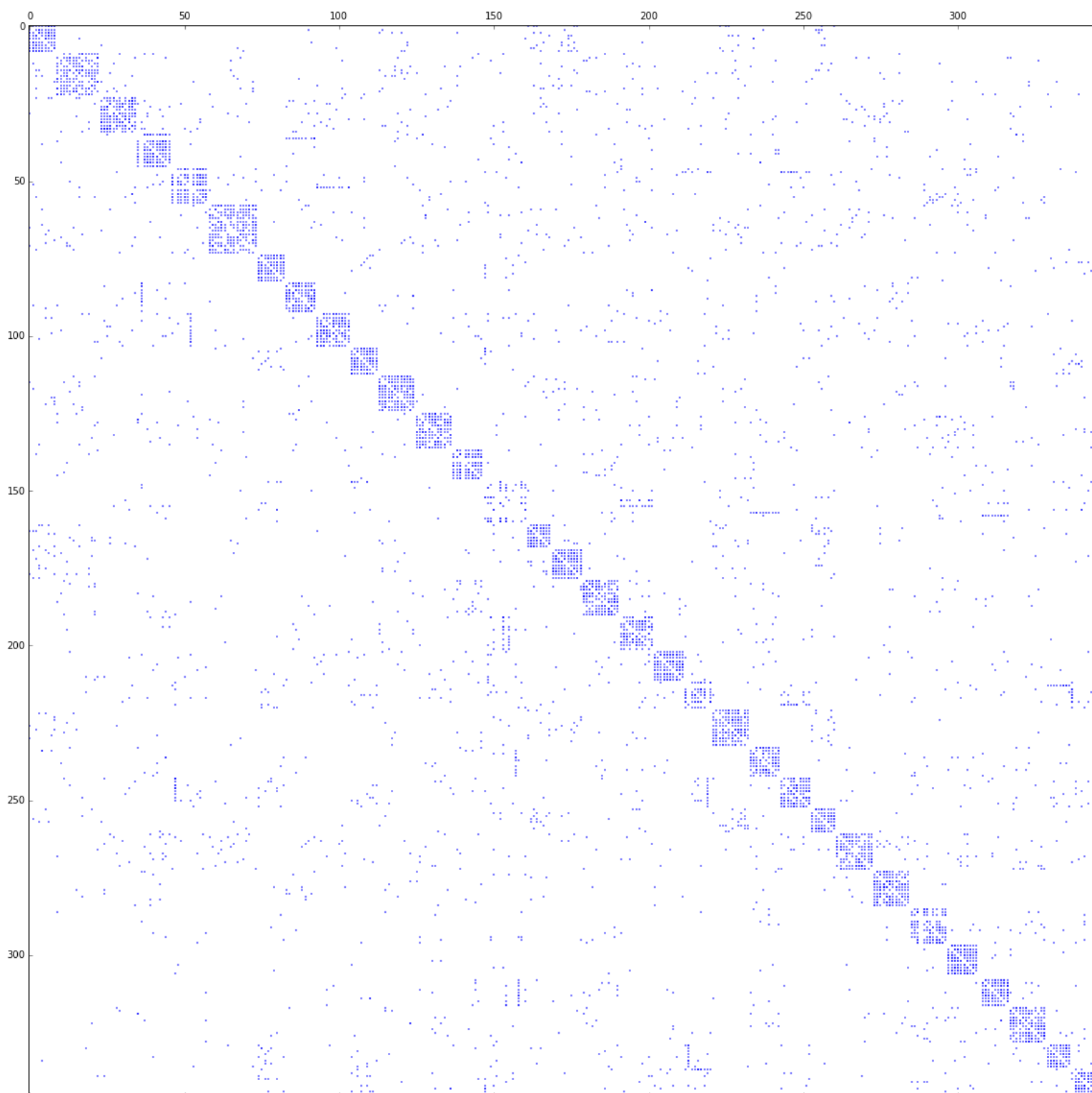
Let's plot the matrix!

In [34]:

```
plt.figure(figsize=(20,20))
plt.spy(winlossmatrix, marker='.', markersize=2)
```

Out[34]:

```
<matplotlib.lines.Line2D at 0x10f9b7828>
```



We'll be using something called the *power method*. Given a random ranking v we'll *play* on that ranking a bunch of times (1000) to see what the outcome would be.

In [44]:

```
A = winlossmatrix

# power method
v = np.random.rand(nteams,)
for i in range(0,1000):
    t=A*v
    v=t/np.linalg.norm(t)
    alph = np.dot(v,A*v)/np.dot(v,v)

I = np.argsort(v)
for t in range(1,nteams+1):
    print('%d.   %s'%(t,teamsName[I[-t]]))
```

1. Kentucky
2. North Carolina
3. Michigan St.
4. N.C. State
5. Ohio St.
6. Syracuse
7. Duke
8. Minnesota
9. Connecticut
10. Seton Hall
11. Notre Dame
12. Missouri
13. Purdue
14. Florida St.
15. Kansas
16. Florida
17. Miami (FL)
18. Virginia Tech
19. Vanderbilt
20. Marquette
21. UCLA
22. Georgia
23. St. John's
24. Mississippi St.
25. Georgetown
26. Saint Louis
27. St. Joseph's
28. Michigan
29. Alabama
30. Maryland
31. California
32. Arizona
33. Cincinnati
34. Pittsburgh
35. Northwestern
36. USC
37. Wisconsin
38. Creighton
39. West Virginia
40. La Salle
41. Gonzaga
42. Virginia
43. Illinois
44. Nebraska
45. Georgia Tech
46. Louisville
47. Tennessee
48. Rutgers
49. Xavier
50. Washington
51. Providence
52. Memphis
53. Colorado

54. Villanova
55. Mississippi
56. UNLV
57. Iowa
58. Baylor
59. Indiana
60. Dayton
61. South Carolina
62. Wichita St.
63. Oregon
64. Boston Coll.
65. Massachusetts
66. Texas
67. Wake Forest
68. Iowa St.
69. Stanford
70. LSU
71. Illinois St.
72. Oregon St.
73. DePaul
74. Clemson
75. Arkansas
76. Drake
77. Temple
78. Kansas St.
79. Auburn
80. Oklahoma
81. Northern Iowa
82. San Diego St.
83. South Florida
84. Oklahoma St.
85. Evansville
86. VCU
87. New Mexico
88. New Mexico St.
89. Missouri St.
90. BYU
91. Rhode Island
92. Southern Ill.
93. Washington St.
94. Bradley
95. Arizona St.
96. Tulsa
97. Texas Tech
98. Penn St.
99. Old Dominion
100. St. Bonaventure
101. Utah
102. Richmond
103. Georgia St.
104. Marshall
105. Duquesne
106. Fairfield

107. UCF
108. Texas A&M
109. Boise St.
110. Geo. Washington
111. George Mason
112. Colorado St.
113. Hofstra
114. Santa Clara
115. Pennsylvania
116. San Francisco
117. Charlotte
118. St. Mary's
119. Drexel
120. Portland
121. Towson
122. Iona
123. Utah St.
124. Nevada
125. Indiana St.
126. Denver
127. Niagara
128. W. Michigan
129. Southern Miss
130. Vermont
131. Davidson
132. West. Kentucky
133. TCU
134. Delaware
135. Butler
136. Stony Brook
137. Northeastern
138. William & Mary
139. Air Force
140. UTEP
141. Ohio
142. Manhattan
143. UAB
144. San Diego
145. Loyola Marymount
146. East Carolina
147. Wyoming
148. Loyola (MD)
149. Belmont
150. Fordham
151. SMU
152. Rice
153. Toledo
154. Fresno St.
155. Charleston
156. James Madison
157. Albany
158. Rider
159. St. Peter's

160. Canisius
161. Oakland
162. Princeton
163. Akron
164. N.C. Asheville
165. Hawaii
166. Brown
167. Siena
168. South Dakota St.
169. Wagner
170. Buffalo
171. Marist
172. Houston
173. Robert Morris
174. Middle Tenn. St.
175. Radford
176. Detroit
177. Oral Roberts
178. Kent St.
179. Tulane
180. St. Francis (NY)
181. Pepperdine
182. Northern Illinois
183. UCSB
184. East. Michigan
185. Hartford
186. UC Riverside
187. Quinnipiac
188. Binghamton
189. Boston U.
190. Murray St.
191. Bowling Green
192. Wis. Milwaukee
193. N.C. Wilmington
194. Long Beach St.
195. Monmouth
196. Central Conn. St.
197. North Dakota St.
198. Arkansas St.
199. Cleveland St.
200. Mercer
201. New Hampshire
202. Dartmouth
203. Ball St.
204. Long Island
205. Holy Cross
206. UMBC
207. Seattle
208. Lehigh
209. Cent. Michigan
210. Bucknell
211. Western Carolina
212. Idaho

213. South Alabama
214. Valparaiso
215. North Texas
216. N.C. Greensboro
217. Mount St. Mary's
218. Fair. Dickinson
219. Yale
220. Coastal Carolina
221. Maine
222. Wright St.
223. Miami (OH)
224. Louisiana Tech
225. Loyola Chicago
226. Lafayette
227. High Point
228. E. Tennessee St.
229. Columbia
230. AR Little Rock
231. Fla Gulf Coast
232. San Jose St.
233. Troy
234. Harvard
235. Wofford
236. Citadel
237. IUPUI
238. Furman
239. UC Davis
240. Samford
241. Florida Intl.
242. Bryant
243. Sacred Heart
244. Florida Atlantic
245. UC Irvine
246. American
247. Presbyterian
248. Cal Poly
249. Montana
250. Appalachian St.
251. Cornell
252. Charleston Sou.
253. Pacific
254. Georgia Southern
255. Mo. Kansas City
256. Austin Peay
257. Norfolk St.
258. Liberty
259. Weber St.
260. Tennessee St.
261. LA Lafayette
262. S.Car. Upstate
263. Northern Colorado
264. Army
265. Winthrop

266. Southern Utah
267. TX Arlington
268. CSU Bakersfield
269. LA Monroe
270. Savannah St.
271. IPFW
272. Northern Arizona
273. Colgate
274. Idaho St.
275. Western Ill.
276. Morgan St.
277. SE Missouri St.
278. Longwood
279. Ill. Chicago
280. Virginia Military
281. CSU Fullerton
282. Florida A&M
283. Lamar
284. South Dakota
285. Morehead St.
286. Navy
287. Campbell
288. Youngstown St.
289. Hampton
290. Jacksonville St.
291. Stetson
292. UTSA
293. Bethune-Cookman
294. Gardner-Webb
295. St. Francis (PA)
296. Lipscomb
297. CSU Northridge
298. TAMU Corpus Christi
299. SIU Edwardsville
300. Jacksonville
301. Utah Valley
302. TX Pan American
303. Chattanooga
304. North Florida
305. Elon
306. Tennessee Tech
307. Sam Houston St.
308. Kennesaw St.
309. Tenn-Martin
310. Delaware St.
311. Sacramento St.
312. Montana St.
313. Eastern Illinois
314. N.J.I.T.
315. Eastern Wash.
316. North Dakota
317. MD Eastern Shore
318. Houston Baptist

- 319. Howard
- 320. Chicago St.
- 321. N.C. A&T
- 322. Texas St.
- 323. N.C. Central
- 324. Coppin St.
- 325. Portland St.
- 326. Stephen F. Austin
- 327. Eastern Kentucky
- 328. S. Carolina St.
- 329. Northwestern St.
- 330. McNeese St.
- 331. Miss. Valley St.
- 332. Central Arkansas
- 333. Texas Southern
- 334. SE Louisiana
- 335. Prairie View A&M
- 336. Southern
- 337. Jackson St.
- 338. AR Pine Bluff
- 339. Grambling St.
- 340. Alabama St.
- 341. Alabama A&M
- 342. Alcorn St.
- 343. New Orleans
- 344. Wis. Green Bay
- 345. Centenary
- 346. Nicholls St.

In []: